CAL ROOM
L LIBRARY
OF MICH.

OCT 30 1936

OCTOBER, 1936

Journal of Educational Research

Registered June 4, 1929

CONTENTS

Complexity of Test Items as a Factor in the Validity of Measurement — *Douglas E. Scates* 77

The Differential Functions of Examinations — *Palmer O. Johnson* 93

Relation between Intelligence and Certain Aspects of Physical Ability — *J. C. Seegers* and *Otto Postpichal* 104

Improvement in Educational Achievement Accompanying a Health Education Program — *Martha Crumpton Hardy* 110

The Computational Arithmetic of Social Experiences of Third Grade Children — *Ebba L. Wahlstrom* 124

Editorial: Training Field Workers and Graduates in Methods of Research — *Carter V. Good* 130

Reviews 133

Research Abstracts and Bibliographies 141

Research News and Communications 146

THE JOURNAL OF EDUCATIONAL RESEARCH is indexed in the Education Index

---

Copyright 1936, by Public School Publishing Co.

Entered as second-class matter at the Post Office at Bloomington, Illinois, under the act of March 3, 1879

JOURNAL OF EDUCATIONAL RESEARCH

EDITORIAL BOARD

A. S. BARR, *Chairman, Professor of Education, University of Wisconsin, Madison, Wisconsin.*

E. J. ASHBAUGH, *Dean, College of Education, Miami University, Oxford, Ohio.* Administration, Supervision, and Teacher Training.

HARRY J. BAKER, *Director, Psychological Clinic, Detroit Public Schools, Detroit, Michigan.* Child Welfare, Guidance, and Development.

LEO J. BRUECKNER, *Professor of Education, University of Minnesota, Minneapolis, Minnesota.* Elementary Education.

CARTER V. GOOD, *Professor of Education, University of Cincinnati, Cincinnati, Ohio.* Teaching, Learning, and Individual Differences.

KAI JENSEN, *Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.* Preschool Education.

CHARLES W. KNUDSEN, *Graduate School of Education, Harvard University, Cambridge, Massachusetts.* Secondary Education.

A. K. LOOMIS, *c/o Board of Education, Shaker Heights, Cleveland, Ohio.* Curriculum Construction.

CLIFFORD WOODY, *Director, Bureau of Educational Reference and Research and Professor of Education, University of Michigan, Ann Arbor, Michigan.* Measurements, Statistics, and Research.

To save time, manuscripts may be sent directly to members of the editorial board as indicated above.

ASSOCIATE EDITORS

FRED C. AYER, *Professor of Educational Administration, University of Texas, Austin, Texas.*

PHILIP A. BOYER, *Director of Educational Research, School District of Philadelphia, Philadelphia, Pennsylvania.*

FOWLER D. BROOKS, *Head of Departments of Education and Psychology, DePauw University, Greencastle, Indiana.*

WILLIAM A. BROWNELL, *Professor of Educational Psychology, Duke University, Durham, North Carolina.*

B. R. BUCKINGHAM, *Lecturer in Education, Harvard University, Cambridge, Massachusetts.*

W. W. CHARTERS, *Director, Bureau of Educational Research and Professor of Education, Ohio State University, Columbus, Ohio.*

VIRGIL E. DICKSON, *Assistant Superintendent of Schools and Director of Bureau of Research and Guidance, Berkeley Public Schools, Berkeley, California.*

EDGAR A. DOLL, *Director, Department of Research, Training School, Vineland, New Jersey.*

WILLIAM S. GRAY, *Director, Teacher Training, University of Chicago, Chicago, Illinois.*

CHARLES E. GREENE, *Assistant Superintendent of Schools, Denver Public Schools, Denver, Colorado.*

M. E. HAGGERTY, *Dean, College of Education and Professor of Education, University of Minnesota, Minneapolis, Minnesota.*

V. A. C. HENMON, *Chairman, Department of Psychology and Professor of Psychology, University of Wisconsin, Madison, Wisconsin.*

KARL J. HOLZINGER, *Professor of Education, University of Chicago, Chicago, Illinois.*

ERNEST HORN, *Director, University Elementary School and Professor of Education, State University of Iowa, Iowa City, Iowa.*

EARL HUDELSON, *Dean, College of Education and Professor of Education, West Virginia University, Morgantown, West Virginia.*

FREDERICK J. KELLY, *Chief, Section of Colleges and Professional Schools, Office of Education, Department of the Interior, Washington, D. C.*

WILLIAM A. McCALL, *Professor of Education, Teachers College, Columbia University, New York City.*

W. J. OSBURN, *Professor of Education, University of Washington, Seattle, Washington.*

SHELTON PHELPS, *President, Winthrop College, Rock Hill, South Carolina.*

DOUGLAS E. SCATES, *Director, Research and Statistics, Cincinnati Public Schools, Cincinnati, Ohio.*

JESSE D. SEARS, *Professor of Education, Stanford University, California.*

FRANCIS T. SPAULDING, *Associate Professor of Education, Harvard University, Cambridge, Massachusetts.*

JOHN L. STENQUIST, *Director, Bureau of Educational Research, Department of Education and Lecturer, Johns Hopkins University, Baltimore, Maryland.*

PERCIVAL M. SYMONDS, *Associate Professor of Education, Teachers College, Columbia University, New York City.*

W. W. THEISEN, *Assistant Superintendent of Schools, Milwaukee, Wisconsin.*

M. R. TRABUE, *Director, Division of Education, Consolidated University of North Carolina, Chapel Hill, North Carolina.*

W. L. UHL, *Dean, School of Education, and Professor of Education, University of Washington, Seattle, Washington.*

CARLETON W. WASHBURNE, *Superintendent of Schools, Winnetka Public Schools, Winnetka, Illinois.*

GUY M. WHIPPLE, *Secretary, National Society for the Study of Education, Clifton, Massachusetts.*

ELIZABETH L. WOODS, *Supervisor, Educational Research and Guidance, Los Angeles City Schools, Los Angeles, California.*

G. A. YOAKAM, *Professor of Education and Director, Courses in Elementary Education, University of Pittsburgh, Pittsburgh, Pennsylvania.*

THE JOURNAL OF EDUCATIONAL RESEARCH is published at Bloomington, Illinois, monthly, except June, July, and August by the Public School Publishing Company. The subscription price is $3.30; $3.50 in Canada or Mexico; and $3.80 in other foreign countries.

Subscriptions and all communications about reprints or other business matters should be addressed to the PUBLIC SCHOOL PUBLISHING COMPANY, at Bloomington, Illinois.

Address all communications on editorial matters to A. S. Barr, Department of Education, University of Wisconsin, Madison, Wisconsin.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume XXX OCTOBER 1936 Number 2

COMPLEXITY OF TEST ITEMS AS A FACTOR IN THE VALIDITY OF MEASUREMENT*

DOUGLAS E. SCATES
Director of School Research, Cincinnati

SOME tests are given for the purpose of measuring rather simple abilities; others are given to measure more elaborate abilities. Whatever the degree of complexity of the ability to be measured, it is desirable that the measuring instrument contain tasks which have a corresponding degree of complexity, if variations in the ability are to be faithfully represented in the measurements. We do not overlook the fact that practical considerations often necessitate the use of some more simple or perhaps more complex form of response which correlates sufficiently well with the ability to be measured so that approximate measurements can be secured. It is well known that many different kinds of performance bear a substantial correlation with each other, so that many tasks of different degrees of complexity are likely to have a substantial relationship. There is, however, considerable loss in such indirect indexing, and the desire for a high degree of validity in measurement usually calls for something more than a dependence upon intercorrelations between abilities. It is the purpose of this paper to examine the demands that validity makes upon the complexity of test items for the measurement of various kinds of ability.

By *complexity* is meant the number of factors which enter into the response to a test item, and the interrelationships which exist between these factors. In the case of a test of mental performance, complexity relates to the thought structure of the response, including perhaps the initial selection of factors which should be responded to. In other words, complexity depends upon the number of significant variables that are required for a satisfactory response, and on the pattern of their interrelationships. We have scarcely as yet arrived at the point where we can say with confidence how many important variables there may be in

* Read before the American Educational Research Association at Cleveland, February 27, 1934.

a response, or exactly what their nature and relationships are, but it may nevertheless be profitable to discuss the matter in general terms. Definite consideration of this aspect of test items may throw a somewhat different light on the construction of "just ordinary" tests as well as standardized tests, and may possibly lead to a somewhat different emphasis in evaluating tests.

Although we are concerned primarily with the character of the response which is made to a test, it becomes necessary for the test maker to turn his attention to the test items and to evaluate them in the light of the kind of response they will normally stimulate. We shall therefore apply the term *complexity* to test items as well as to responses—meaning that the test item must be appropriately complex to stimulate a response of the desired complexity, in so far as the test item is depended upon as the significant stimulus. While the matter of appropriate complexity of test items is one that is not ordinarily ignored entirely, it is typically dealt with only incidentally, and not directly. That is, the test maker commonly selects a particular bit of "knowledge" and casts it into a form which he deems appropriate for an item in his test. The complexity which the item may have is determined mainly by the character of the test which is being constructed, and not by careful attention to this particular aspect of the item. It is more or less taken for granted that all items of the same type are of the same complexity.

If the test is being made by the research worker, he will of course analyze his test items statistically, but that is likely to lead only to the casting out of items which do not seem to be as good as others; it does not improve those which are left. Hence, the test in its final form is not likely to be appreciably better than a group of the best items as originally constructed. It is therefore worth while to give some thought to means by which the quality of test items may be improved at the outset, and the giving of definite attention to the matter of complexity is suggested as a supplement to the present general emphasis upon selection of items. Moreover, in the case of the common teacher-made test, such an emphasis is particularly important, for in these tests items are usually not even selected.

The general concept of complexity of measurement will be developed first by reference to simple situations in the field of physical measurement, where the relationships can be readily seen, and subsequently by calling attention to its application in the field of educational measurement. Turning our attention to a simple illustration from the physical

field, we may note the difference between linear measurement, square measurement, and cubic measurement. These are three different kinds of measurement because they differ in the number of factors (planes, dimensions, or variables) which are included in them. In fact, even though they may all be expressed as functions of distance, one system of measurement may not be transformed into the other without the addition or elimination of one or more factors. When a seventh grade girl recently asked her teacher, "How many square feet are there in a cubic foot?" she failed to recognize that differences in the degree of complexity of the types of measurement involved rendered her question unanswerable. Measurements having different degrees of complexity (that is, representing different numbers of factors, such as linear, square, and cubic measure) are regarded as constituting fundamentally different systems of mensuration.

It will be noted at once that any one of these types of measurement is not appropriate for measuring outside of its own field. Linear measurement will not suffice for area, and square measurement will not suffice for volume. The content of a two-foot cube is not two or four times that of a one-foot cube, but eight times. It is of course true that we are sometimes satisfied with measurements that are deficient in some of the factors which we desire to be present. For example, we might be satisfied to measure only the length of a number of boxes, when we were really interested in their volume, though it is obvious that we would be contenting ourselves with very rough suggestions as to actual or relative volume. We commonly are content to indicate the length of a book by the number of pages which it contains, knowing full well that the number of words on a page will vary with the size of the page, the size of the type, and several other factors. We may be satisfied, as a first indication, with knowing the number of rooms in a house; but we usually want to follow this crude type of measurement with information concerning the area of the rooms, their shape, and the pattern in which they are arranged to make up the house. Just how closely the complexity of the measurement need follow the complexity of the character which it is desired to measure is a matter which is in all cases dependent upon one's purpose; but it may be suggested as a general principle that it is scarcely safe to ignore factors unless one knows pretty well their effect upon the significance of the measurements which he obtains.

It is perhaps desirable to call attention at this point to the distinction between relevant and irrelevant factors in measurement. All measure-

ment is of course highly selective, disregarding many aspects of the objects which are not desired to be included in the measurement. In the case of boxes, for example, if one is interested in measuring volume, he will not include in his measurements such factors as the kind of wood, its color, its thickness, its moment of rupture, its specific gravity, etc. Excluding such irrelevant characteristics, however, is a different matter from excluding factors which constitute essential elements in the general character which one is seeking to measure. In measuring volume one cannot normally omit with impunity any of the three relevant factors of length, breadth, or depth. Whether a factor is relevant or not is of course dependent upon one's purpose—we may be interested in thickness of wood rather than in volume. In considering the complexity of measurement which is appropriate, therefore, one needs to distinguish between those factors which are essential to his purpose and those which are irrelevant.

It is not always easy, even in the field of physical measurements, to identify and provide for all of the factors which should be measured. A somewhat more complex illustration may serve to show this. Some years ago the statement was made that people would help avoid traffic congestion if they would ride street cars. The argument advanced was that when the length of a street car was divided by the number of people it would hold each person was found to require only about one-fifteenth as many feet as he did in an automobile. Without any immediate interest in the correctness of the statement, it might be pointed out that the comparison which was used ignored at least four important factors which should be taken into consideration before the figure could take on real significance. First, square feet, or perhaps "lane-feet," should be used instead of linear feet, since many highways provide for more than one traffic lane. Second, the linear distance which must be preserved *between* vehicles is perhaps more important even than the length of the vehicle itself. Third, the comparison used the maximum load for a street car and the minimum load for an automobile; the average load for both, at the rush period, should have been used. Fourth, congestion is in part a matter of average rate of travel, taking stops into consideration. There are other less important factors. The figure used, representing only linear feet per person, could not have much significance when a figure giving lane-feet per person per mile per hour, on the average, contained the minimum number of important factors.

The mathematically minded person would say that the unit required in this problem was multi-dimensional. This is not a common term in education, but it is a suggestive one. It means that, instead of being expressible in terms of two or three dimensions, such as area and volume are, the nature of the thing to be measured is so complex as to require many dimensions, or factors, in the measurement. It is somewhat difficult to extend one's thinking into dimensions beyond the third, and this is of course only a mathematical figure of speech. It may, however, be suggestive to think of a jewel cut with many irregular facets, each one having its own peculiar shape. One can readily see that only a very imperfect description of the stone could be obtained by viewing it from any single point. To reveal the true shape of each facet would require a plane parallel to it, making as many planes as there were sides. If each such plane or dimension is thought of as representing a separate variable in the total description of the stone, the need for many of them can be appreciated.

The foregoing illustrations have been drawn from the physical fields in which the appropriateness of a certain degree of complexity in the measurement is obvious. But it should be equally clear that in psychological fields measurement must also be appropriately complex. In human behavior we are likely to find more significant factors, and a more complicated interrelation of these factors, than we are in physical fields. The factors in psychological work are also less generally recognizable, so that it is difficult to ascertain just what ones should be included. It remains certain, however, whatever the difficulty involved in securing satisfactory measurement, if we attempt to measure complicated psychological response tendencies with measuring instruments which are either too simple or too complex for our particular purpose, we cannot expect to get a picture that will reveal the true conditions of those characters which we desire to study.

To attempt a list of the various factors which may condition a human response is not within the scope of this paper. We shall, however, refer to three general groups of factors for purposes of illustrating different degrees of complexity in psychological measurement. We may refer to these three groups simply as intellectual, environmental, and personal. In the first group, the intellectual factors, will be included the abilities and the types which are commonly tested in our ordinary tests. If one thinks in terms of objective tests, including most of the standardized tests commercially available, he will think chiefly of knowledge of facts,

including a certain proportion of relationships of various kinds, most of which have been studied as factual relationships. The informal teacher-made tests of the objective type also abound in this type of item.

It would be profitable if teachers and other test makers would undertake a psychological analysis of the complexity of the processes required to make a proper response to the more common test items, and compare their findings with the complexity of the types of mental activity which they commonly assert they are teaching to develop. It seems probable that we are on the whole attempting to measure very complex types of intellectual learning with very simple tasks. For example, we ascertain whether a pupil can respond to a "knowledge" item without commonly expressing any interest or concern over what this knowledge may mean to the pupil. There is an infinite variety of kinds and degrees of relationships between specific response tendencies (such as "knowledge") and the background of thinking and use into which they fit. The measurement of these more complex processes calls for a many-factored complexity within the test items which are used. Even in those schools which are content to have their teaching follow along the traditional line of emphasis almost exclusively upon intellectual aspects of learning, there is still a great opportunity to improve the type of testing which is commonly done.

Countless tests that are given to measure intellectual achievement consist of rather simple recall or identification items of various kinds solely because it has become fashionable to give tests of this character and not because anyone can justify them as appropriate measures of the objectives of education to which most teachers will at least give lip service. The mental processes commonly measured are of the very simplest kind, and are of relatively superficial significance as indicators of real growth when education is looked at in the large. One should examine the objective tests which he has used during the past year to see how many items in the course of a year have actually required the ability to make generalizations, to apply general principles, to draw inferences, to detect similarities and differences, to evaluate, to appreciate, to detect cause and effect relationships, to marshal facts, to support a position, etc.,—in areas where the facts and relationships called for have not been taught as such. Or if one feels that an adequate understanding of a complex subject is an important intellectual ability even when directly taught, how many of our tests, and particularly the standardized tests in the Social Studies—make any apparent attempt to reflect the

pupil's knowledge of why facts are as they are? A real understanding would call for a knowledge of what factors have caused conditions to be so, and what the relative importance of these factors is. As contrasted with this, the majority of test items seek to ascertain only whether the pupil can recall some fact, which has previously been taught directly as a fact.

As stated earlier, for certain purposes of measurement one may be content to ignore those factors which cause one kind of ability to be different from another kind; but when one does so he will be careful not to draw conclusions of large significance. It is somewhat difficult to believe that the conclusions which teachers must draw from their measurements can be classed as not of large significance. It is highly dangerous to assume that we are adequately indexing intellectual achievement of any large moment when we use test items of the simpler kinds. If one is interested in testing the significance of a fact in the pupil's thinking, in ascertaining how many other ideas a fact is associated with in the pupil's mind and how it fits into a large pattern of his responses, in testing the pupil's ability to follow through an appropriate sequence of ideas in order to reach a conclusion, in getting some insight into the pupil's ability to draw upon a large and varied background in order to make proper evaluations, in looking into the pupil's ability to see relationships quickly in new material, and in studying many other mental operations which are essential to the development of real mental power, then one will have to discard the vast array of highly mechanical, narrow test items which generally encumber educational tests at the present time, and seek out forms of test items which are appropriate to one's purpose.

We are, however, interested in more dimensions or aspects of performance than those which are represented by intellectual ability. We are interested in some of the factors which cause this ability to vary, or take on special forms. We are, for example, all aware from personal experience that we can recall something at one time and not at another; we can think of things to say, or we can work out solutions to problems, under one set of circumstances and not under others. We have gradually come to realize that even such an apparently simple thing as an item of knowledge is not a single thing, nor is it a constant thing; its existence is contingent upon the strength of a tendency for the individual to respond in that particular way, when various conditions are stimulating responses of other kinds. We are therefore interested in an item of

knowledge, or some other ability, not only under one set of conditions but under a number of different sets of conditions.

In this second group of factors, referred to as environmental, we should include those kinds of stimulation which arise out of the physical conditions, social conditions, and perhaps the emotional set, involving the stimulation of effort, various distractions, and confusions. The test may be presented in such a way that all, or most of the group will respond with a great intensity of effort, or they may engage in their tasks with only an indifferent attitude. There may be various physical distractions, such as noise, heat, cold, etc. The test administrator may emphasize speed or caution, guessing or certainty, or other general attitudes. It is clear that there may be considerable stimulation of a significant kind which does not arise from the test items themselves. Certain factors which are present in the responses may arise from the general conditions of the testing. As such, they fall somewhat outside of the present discussion of test items though essentially they represent an extension of the principle being discussed. It may be profitable to give some attention to them in passing.

Especially important in this second group of factors which condition responses are certain types of social stimuli, particularly those which enter into social performance. As teachers, we sometimes fail to recognize the differences which obtain between responses under formal classroom conditions and responses under more normal social conditions. For example, it is a rather common experience for one to find that his arithmetic does not function as well under typical conditions of life as it does in the classroom. When, for example, one is buying potatoes in the grocery store priced at twenty-five cents for fifteen pounds, and one wants only fifteen cents worth, he may want to know how many pounds the scales should show in order for him to get his share. He knows that the problem is simple, if he were only back in school, or back home; but it is not simple when he must hurry to figure it out before the grocer takes the sack off the scales, and when he must have enough confidence in his answer so that he will be willing to raise a question about the weight if it does not agree with his mental calculations. There are many people who can do their formal arithmetic well, who wait until they get home from the store to see whether the storekeeper gave them the right change or not, because they get confused in trying to count change in the social situation and are not willing to expose their confusion by taking a little more time.

The trouble in all such cases, where the abstract skills have been mastered, is not that the person's knowledge of academic arithmetic is faulty, but simply that it is academic. If we assume that we are concerned with the development of functional skills under practical conditions, the test for such learning must be something more than the routine sort. Probably the test situation here becomes more important than the particular test item (if the "knowledge" involved may be thought of as an item). The element of importance again is the complexity of the total stimulation pattern, and it is rather difficult to distinguish that part of the stimulation which should be thought of as originating primarily in something corresponding to a particular test item, from that part of the stimulation which should be thought of as growing out of the general situation. The point of trying to make such a distinction is of course a theoretical one only; the practical significance of the illustrations is that it is often essential to secure a complexity of stimulation that is far beyond that usually attained in our testing, if we are to measure the most serviceable aspects of some of our learning.

To achieve this complexity of stimulation not infrequently calls for something other than paper and pencil tests. Consider for example, the ability to speak informally before an audience, involving as it does a selection of social stimuli, a disregard for many of the stimuli, and a proper relationship between attention to selected aspects of the social situation and attention to one's purpose in speaking—all of which are in addition to the technical or more intellectual phases of the talk, such as having something of significance to say, having it well organized, having a command of correct English usage, etc. These more general, or perhaps non-academic, aspects of the individual's total adaptation, and their interplay with the more formal, more academic aspects obviously represent response complexities which are far beyond the scope of typical tests to deal with. They are, however, none the less important as educational outcomes; in fact, it may be held that if the more narrow intellectual forms of learning do not develop into the larger abilities which are of common importance in life at large there is some question about the value of teaching the narrow forms of learning. At any rate, it will be apparent that the total stimulation pattern which obtains in everyday affairs is likely to be very different from that which surrounds the giving of routine tests in the classroom. It may be worth while pausing to consider the practical validity of our measurements when based upon the

knowledge, or some other ability, not only under one set of conditio but under a number of different sets of conditions.

In this second group of factors, referred to as environmental, should include those kinds of stimulation which arise out of the physi conditions, social conditions, and perhaps the emotional set, involving stimulation of effort, various distractions, and confusions. The test m be presented in such a way that all, or most of the group will respond w a great intensity of effort, or they may engage in their tasks with an indifferent attitude. There may be various physical distractions, su as noise, heat, cold, etc. The test administrator may emphasize speed caution, guessing or certainty, or other general attitudes. It is clear t there may be considerable stimulation of a significant kind which does arise from the test items themselves. Certain factors which are pres in the responses may arise from the general conditions of the testing. A such, they fall somewhat outside of the present discussion of test ite though essentially they represent an extension of the principle being di cussed. It may be profitable to give some attention to them in passi

Especially important in this second group of factors which conditi responses are certain types of social stimuli, particularly those wh enter into social performance. As teachers, we sometimes fail to rec nize the differences which obtain between responses under formal cla room conditions and responses under more normal social conditions. F example, it is a rather common experience for one to find that his arit metic does not function as well under typical conditions of life as it d in the classroom. When, for example, one is buying potatoes in t grocery store priced at twenty-five cents for fifteen pounds, and one wan only fifteen cents worth, he may want to know how many pounds t scales should show in order for him to get his share. He knows that t problem is simple, if he were only back in school, or back home; but i is not simple when he must hurry to figure it out before the grocer tak the sack off the scales, and when he must have enough confidence in h answer so that he will be willing to raise a question about the weight it does not agree with his mental calculations. There are many peo who can do their formal arithmetic well, who wait until they get hom from the store to see whether the storekeeper gave them the right chan or not, because they get confused in trying to count change in the soc situation and are not willing to expose their confusion by taking a lit more time.

The trouble in all such cases, where the abstract skills have been mastered, is not that the person's knowledge of academic arithmetic is faulty, but simply that it is academic. If we assume that we are concerned with the development of functional skills under practical conditions, the test for such learning must be something more than the routine sort. Probably the test situation here becomes more important than the particular test item (if the "knowledge" involved may be thought of as an item). The element of importance again is the complexity of the total stimulation pattern, and it is rather difficult to distinguish that part of the stimulation which should be thought of as originating primarily in something corresponding to a particular test item, from that part of the stimulation which should be thought of as growing out of the general situation. The point of trying to make such a distinction is of course a theoretical one only; the practical significance of the illustrations is that it is often essential to secure a complexity of stimulation that is far beyond that usually attained in our testing, if we are to measure the most serviceable aspects of some of our learning.

To achieve this complexity of stimulation not infrequently calls for something other than paper and pencil tests. Consider for example, the ability to speak informally before an audience, involving as it does a selection of social stimuli, a disregard for many of the stimuli, and a proper relationship between attention to selected aspects of the social situation and attention to one's purpose in speaking—all of which are in addition to the technical or more intellectual phases of the talk, such as having something of significance to say, having it well organized, having a command of correct English usage, etc. These more general, or perhaps non-academic, aspects of the individual's total adaptation, and their interplay with the more formal, more academic aspects obviously represent response complexities which are far beyond the scope of typical tests to deal with. They are, however, none the less important as educational outcomes; in fact, it may be held that if the more narrow intellectual forms of learning do not develop into the larger abilities which are of common importance in life at large there is some question about the value of teaching the narrow forms of learning. At any rate, it will be apparent that the total stimulation pattern which obtains in everyday affairs is likely to be very different from that which surrounds the giving of routine tests in the classroom. It may be worth while pausing to consider the practical validity of our measurements when based upon the

relatively artificial conditions affecting responses while pupils are sitting in their seats writing, as contrasted with being out in the world doing

The third group of factors which condition performance has been referred to as personal. In this group are included such aspects of the individual's larger habits of life as purposes, interests, working standards of performance, emotionalized attitudes, and scales of value or preferences. These lie to a certain extent in the fields of personality and character; but they also have a close relation to the intellectual aspects of learning. In fact, it has been suggested that pupils study the common school subjects primarily in order to give teachers an opportunity to develop these orienting and dynamic aspects of the individual. Whatever their relative importance, there will be little question that they are important. While more technical forms of knowledge may greatly raise our material standard of living, it is the status of the traits in this third group which determines the essential soundness of our society. Almost any school principal, educational philosopher, or commencement speaker will declare that these are the forms of learning for which schools really exist; yet they are seldom measured at all—as though they were immaterial.

There is some truth in the suggestion that we know comparatively little at the present time about the measurement of personality and character—aside from the use of somewhat questionable paper and pencil tests. It is also true that an adequate appraisal of many of these traits must rest upon observation rather than formal measurement. But the facts in the case scarcely justify present practices—or lack of practice. Certainly it is possible to gather systematic information concerning pupils' interests, concerning their use of leisure time, concerning their preferences, and other tendencies, both by blanks which they fill out, and by observation. Further, if these traits are really as important as most educators say they are, it would be worth while to reorganize the school curriculum so that pupils are given a large opportunity to exercise these characteristics, develop them, and exhibit them. If this is done, there will be some opportunity for gathering relatively significant data on the extent to which the rather general or background response tendencies are present in each pupil.

We are not here concerned, however, with making constructive suggestions; we are concerned only with pointing out that the usual tests which are given in school leave untouched this vast field of learning which is of fundamental importance. Typical test items are simply inadequate

to deal with these larger response patterns, many of which are working habits rather than intellectualized habits. It seems probable that ingenious tests may make further progress in this field, as some already have done; but for the present we must look upon our routine achievement tests—which constitute the great bulk of the testing that is now done—as representing a very small segment of the total learning which we are striving to cultivate. They simply do not possess the complexity appropriate to stimulate those response factors which we are interested in when we look at life in the large.

We may turn our attention from these several groups of factors which contribute to varying degrees of complexity in response tendencies, to consider several questions which arise with reference to the complexity of test items. Perhaps the question first to arise concerns the justification of narrow or specific measurements for certain purposes. It must of course be recognized that there are purposes in education which call for the measuring of less complex abilities. One may, for example, want to measure speed of reading, vocabulary knowledge, word recognition, the ability to get the general meaning of a paragraph, etc., as separate and specific aspects of reading. Such is usually the case where we are interested in diagnosis, and one is interested in such aspects for purposes of guiding instructional emphasis. It is true also that recent scientific work in measurement has been directed toward the securing of more specific, more simple, measuring units.

Two things should be said about measuring simple educational outcomes. In the first place, it is unfortunate for any teacher to become so attentive to the narrower aspects of a subject that she fails to give adequate emphasis to the larger outcomes in that field. In reading, for example, it is of great importance that the child from the very start make progress not only in the mechanics of reading, where tests are numerous, but also in such phases as good taste in his preferences for material to read, appreciation or enjoyment in reading, the habit of reading for pleasure and for intellectual gain, a concept of a balanced program of personal reading, and the like. Such considerations lessen somewhat the *relative* importance of narrow measurements. Without any thought of disparaging such measures when properly used, it is a matter of common observation that they are grossly overemphasized, and when so used they may direct the teacher's attention away from the larger outcomes, and thus serve a very questionable purpose.

In the second place, it should not be lost sight of that the insig provided by narrow measurement becomes much more valuable wh adequate *general* measures of the ability as a whole are available, to affo a picture of the pattern of articulation between the detailed aspects. Th is, it is possible in teaching to emphasize one aspect and then another the general ability, as specific tests reveal that one and another aspe is weak, and when all of these have been given due attention, we ma still have a learning product that is wholly unsatisfactory. We ma never lose sight of the fact that a particular ability calls for an inte relation of many aspects which is just as important in a mental or soci ability as coördination is in muscular performance. Mechanical thing may be taken apart and put together again more or less at will; growt products on the other hand eventuate over a period of time as the resu of a continuous process of inter-adjustment between the various signi cant factors. Such products are too complex to be adequately dealt wi in terms of their separate aspects, for the element of structural patte is as important in a growth product as are the different individu elements which enter into this pattern. It follows that there is a pos bility that the development of separate aspects of a skill seriatim (in s far as such might be possible) might result in a net educational los In so far as we have appropriate *general* tests for our school subject in which the tests call for appropriately complex responses, these ca be used as measures of real progress, to keep the teacher from thinkin too much in terms of piecemeal instruction devoted to one aspect of subject at a time. Teaching is a very delicate art which may easily b thrown off balance by testing which is not well oriented.

One may raise the question as to why it is not better to secure measure of a total ability through putting together detailed measureme of its many aspects. This is practically another form of the question ju dealt with. The structural pattern of the interrelation among ma specific aspects may be much more important than one or all of th aspects taken individually. For example, it would be quite possible fo an individual to have all of the supporting skills that enter into publ speaking well developed and still be almost a complete failure as a speake Of course it is not likely that one would develop these aspects in isolatio but neither is it likely that he will, under normal school conditions, de velop them in proper relationship. As a very simple instance, we ma note that people have on the whole made those general adaptations t their environment which enable them to manage their hands without co

serious attention; yet how many of these people can stand before a sizeable group without being embarrassed by the fact that their hands are receiving far too much of their own attention and that they cannot find anything appropriate to do with them? The skills, or adaptations, which we have made in one set of circumstances, are readily disrupted and rendered useless by some slight change in one's relation to his environment—such as by moving from one's seat in a group to a position in front of the group.

To talk about developing each of the aspects of an ability separately, or measuring them separately (which often leads to an attempt to develop them separately) is to overlook entirely the fact that an ability is the result of an adaptive growth, and comes about through the establishment of many interrelations that are of the greatest importance. We may be able to discover deficiencies in certain aspects of an ability by specific measurement but we cannot by such measurement discover the deficiencies which may exist in the relationships between the aspects. It is not possible to develop such aspects as formal skills, adaptation to situations, and adaptation to purpose, in properly integrated form, by developing each of these aspects separately, and if they are measured separately, we do not know whether they fit together into a desirable general pattern or not. Effective speaking, effective writing, effective thinking, and many other complex abilities, including perceptual skills, have composite aspects which cannot in any way be touched by considering simple analyzable aspects separately.

There is a further question that may be raised. That is the question of whether complexity of measurement should not be sought in the test as a whole, rather than in the items individually. It may be argued, for example, that a test should consist of a good many items on each topic, and that if these items are varied in form the test as a whole will reflect many aspects of knowledge about the topic and hence will achieve the degree of complexity desired. This question must be considered in several steps. In the first place, it must be recognized that a simple increase in the number of items does not in itself increase the complexity of the total measurement, at least in the sense here being discussed. We may increase the thoroughness with which we explore the knowledge of any topic by using more test items; but if these items are homogeneous we do not thereby increase the number of factors represented in the total measurement except in the sense that each separate mental stimulus may be said to be a different factor.

It is, as it were, taking more measurements in the same plane, rather than increasing the number of planes or dimensions which are represented in the measurement. Referring to a physical example, it is like measuring only the height of more and more boxes, when one may really desire to know the volume of the boxes. In this case the number of factors is increased only in the sense that we think of some different factor as causing each box to be a different height; but clearly no number of measurements of height, however great, would ever give us a measure of the volume of the boxes. Similarly, if we are using test items of the relatively simple knowledge type, no increase in the number of them on a chapter in a history textbook will give a picture of the pupil's ability to utilize his knowledge in drawing practical conclusions about the affairs of today through applying the possible lessons which can be learned from the history chapter. A test on upper elementary grade reading which is based on a knowledge of the authors or characters in various pieces of literature, might contain an infinite number of items and still afford no reliable information concerning the ability of the pupil to read literature with considerable enjoyment—nor concerning his disposition to do so. The complexity of a test is determined by the complexity of its individual tasks, and their complexity is not increased through the mere addition of more items.

If we now turn to another aspect of the question, and consider the effect of including in the test items of different degrees of complexity, we note that this produces something in the nature of a scale, the items of different complexity tending to vary in degree of difficulty. This introduces a problem in knowing how to interpret the score. Presumably the test will deal with a number of topics, and the instructor will include items of varying complexity on each topic. The simple sum of the satisfactory responses on these different kinds of test items for each topic is not quite as understandable as would be the score on a scaled test which was homogeneous—that is, concerned with only one topic or one ability. Drawing again upon the analogy of boxes, the total score is like ascertaining the total number of boxes that are satisfactory with reference to various aspects, some boxes being measured for length, some for area of a particular side, and others for volume. Granting that for certain purposes we might be interested in these separate aspects of boxes, nevertheless the total number of "satisfactory" boxes arrived at by this process of measuring one aspect of one box and another aspect of another box does not tell us how many are satisfactory with respect to each one

of the several criteria, such as length, area, or volume—and for any reasonable purpose we should want to know the number of boxes that were satisfactory for each separate aspect. Further, it would appear that even the latter type of knowledge would not generally be satisfactory for educational purposes. It would appear that the mixture of items on a test with varying degrees of complexity, if the items are really different enough to cause them to measure different kinds of mental activity, is more of a handicap than a benefit. In the interests of significant measurement, the items should either be put on separate tests, each having a homogeneous level of complexity, or at least the items should be grouped according to complexity and each group scored separately as representing different aspects of the ability. Items having a considerable variety of complexity are of course not necessary in order to achieve any desired degree of complexity in the test, since the maximum complexity of a test is that of its most complex items. This complexity is not to be sought through the simple combination of responses on different items, which must fail to reveal the pattern of relationship between the factors which might be sampled.

In summary, a test which consists of items that are more simple or more complex than the response tendencies which are to be measured must result in some loss of fidelity in measurement. In the field of physical measurement, appropriate complexity is accepted as axiomatic, and measurements which include different numbers of variables, or which represent the variables in different combination patterns, are regarded as comprising fundamentally different systems of mensuration. Units of measurement which have different degrees of complexity simply measure different things. In the measurement of educational attainment we have differentiated between different fields of learning but we have not generally made corresponding distinctions of a psychological sort between the types of ability called for by test items with different degrees of complexity.

It seems probable, furthermore, that in many cases we have not well gaged the complexity of our test items to the complexity of the abilities which we desire to measure. The common run of test items are almost purely knowledge items. If, on the other hand, one studies various objectives of instruction, one is prompted to raise a number of questions. Are we not to be interested in what a bit of "knowledge" means to a pupil, and how well it will serve his purposes? Are we not concerned with the extent to which the necessary interrelations between this "knowledge"

and appropriate forms of expression for its use in practical purposes have been set up? Have we no responsibility for seeing that we have fitted this "knowledge" into a dynamic scheme of thinking on the part of the pupil so that he can draw upon it as a basis for evaluation and for constructive thinking? Are we not at all concerned with the extent to which the pupil has developed the capacity to strip himself of the accustomed classroom correlatives and demonstrate the ability to respond properly under the various conditions of normal life with their peculiar distractions and with the various pressures of different purposes? True, some of these things carry us outside the range of formal tests; but it is worth something to recognize this fact so that we will not continue to rely upon our common tests as adequate measures of everything we need know.

Perhaps the chief source of difficulty in our current measuring instruments is that test makers have been content to accept for their guidance the relatively superficial concepts of educational objectives which have widely prevailed in practice. A very mechanical and superficial type of measurement has naturally followed in the wake of a generally inadequate concept of the responsibilities of common education. But test constructors, including teachers who make their own tests, must recognize the very strong influence which tests have upon instruction, and must evaluate tests at least in part on this basis. Test makers at the present time bear considerable responsibility for producing tests which will aid in giving an increased emphasis to the more fundamental purposes of education. It is the premise of this paper that one step in this direction is the examination of test items with reference to the appropriateness of their psychological complexity for dealing with the more important educational outcomes. The time has come for us to recognize that we cannot adequately measure learning products of great complexity with test items that are designed to reflect only the most simple kinds of mental activity.

THE DIFFERENTIAL FUNCTIONS OF EXAMINATIONS

PALMER O. JOHNSON
University of Minnesota

THE construction of examinations is a part of the normal occupation of almost every teacher. With this as with all accepted practices it is necessary that the question should be asked from time to time, What is the purpose of the examination and how well is the purpose accomplished? If the purpose is to obtain an appraisal of student accomplishment then it is highly important to make the measurement as accurate as possible. This point of view directs one's attentions to the aims of instruction. If examinations are to give measures of the extent to which each of the purposes of instruction is attained, then all of these purposes that are amenable to measurement should be represented in the examination. The fulfillment of this need for measures of differential functions is still remote with respect to many of the purported outcomes of instruction. Even for those outcomes more susceptible at present to measurement the construction of a valid examination is not a simple matter either in theory or in practice. At least three qualifications appear fundamental for those who are to construct good measuring instruments: (1) thorough knowledge of the field, (2) the ability to resolve the subject into parts proper to testing, and (3) functional knowledge of the principles of test construction. These attributes are not commonly found in the same individual. With the view of improving present examination practices, the University of Minnesota Committee on Educational Research has a program in operation in which the instructors of certain courses, graduate assistants, and examination counsellors are coöperating. It is not the purpose here to describe this plan in operation. However, since all participants in the examination program are mentioned in the central problem, a description of some of their activities is essential. These will be pointed out in connection with the discussion of the main problem which concerns itself with:

1. The measurement of some purported outcomes of instruction.
2. The interrelations of the abilities measured.
3. The separable and combined importance of certain variables.

The instructor is the individual to determine the materials of instruction and the methods of presentation, and to define the outcomes which he hopes to attain through their administration. The primary function of the counsellor is to develop techniques of test construction which will

reveal the extent to which the outcomes have been attained. In order for the counsellor to do effective work it is highly desirable that he have a background in the field in which examinations are to be developed.

While some work has been done upon measures of outcomes other than those reported here, this discussion is restricted to certain aims which have been stated in terms of what the student is expected to do with the materials of instruction. These aims or outcomes are:

1. The acquisition of the terminology or vocabulary distinctive to the subject.
2. The acquisition of fundamental principles and other factual information.
3. The ability to apply the knowledge of vocabulary, principles, and other facts in the interpretation of situations and problems.

In order to secure an authoritative and specific record of the materials of instruction, graduate assistants were used. The assistant was selected largely on the basis of the recommendation of the instructor or department with respect to his knowledge of the field in which the examinations were to be constructed. He attended the class meetings where he made a complete record of the materials presented, studied the basic readings, and explored other sources which might be of value in developing test situations. These materials were then analysed and classified into technical terms, principles, and other factual information. Such were the sources for the content of those examinations designated later as examinations on terminology and examinations on principles and other factual information.

The examination purporting to measure the ability of the student to draw upon his knowledge of vocabulary, other factual information, and principles requires further description. It is recognized that the tests of such outcomes so far developed are empirical tests constructed and applied with incomplete knowledge of the abilities measured. The efforts in this direction appear justifiable, however, on the basis of the importance this objective holds among those advanced as aims of the educational process. An objection to an examination that tests only the acquisition of vocabulary and other factual information is the illusion on the part of the student who feels that if he is able to obtain the answer the teacher desires, he has mastered the difficulties of the subject. A knowledge of vocabulary and fundamental facts or principles is not the end in itself but rather the great trunk from which the spreading branches spring. In

order to motivate the student toward the development of the ability to apply knowledge the measurement of the extent of this acquisition should also be a part of an examination.

As bases for determining the validity of those test items designed to measure the ability of the student to draw upon his knowledge of facts or principles in the interpretation of problems or situations the following means were employed:

1. Determining whether or not the test situations were new. The method of obtaining a record of the materials of instruction, described above, was relied upon in deciding whether or not the test situations had been interpreted by instructor or author. This analysis also enabled the detection of whether or not the basic principle or fact underlying interpretation had been included in the course content. Current research and other scientific literature provided content for a considerable number of the situations; others were improvised.[1]
2. Submittal of the test situation both previous and subsequent to instruction. If the student can interpret the problem before instruction, obviously this performance cannot be attributed to instruction in the course. This is a desirable criterion of validity but could be applied to only some of the test items during the first year of the experiment.
3. Judgment of scholars in the field as to the correct interpretation of the test situations. The instructors contributed substantially by passing judgment on the correct interpretation of problem situations from the standpoint of fact and whether or not the ability to interpret called for a knowledge of the principle considered basic to the solution. They also contributed a number of the test situations.

Methods (1) and (3) were applied in all cases; method (2) to a sampling of the situations. In addition some experimentation was conducted in comparing the free responses with the controlled responses of the students.

An illustration of the items used in the examinations (1) on vocabulary, (2) on the acquisition of facts or principles, and (3) items calling for the interpretation of a problematic situation in the field of physics is:

[1] It is possible but not highly probable that students encountered some of these sources. The large number of test situations, in all cases well over a hundred, would render it extremely unlikely that any significant number would be met with. It is not common practice for undergraduate students to read current research or widely beyond course assignments.

(1) Part I—Vocabulary

1. angle of incidence
2. angle of minimum deviation
3. angle of refraction
4. critical angle
5. diffraction
6. interference
7. mirage
8. parallax
9. polarization
10. real image
11. reflection
12. virtual image

1. ( ) Smallest angle between incident and emergent ray of a prism.
2. ( ) Change in direction of a ray of light in passing from one medium to another.
3. ( ) Change in the direction of propagation of light in a given medium after encountering the surface of another medium.
4. ( ) Image through which light actually passes.
5. ( ) Angle between a perpendicular to a surface and the ray of light striking the foot of the perpendicular.
6. ( ) Angle between a perpendicular to a surface and a ray of light after passing through the surface at the foot of the perpendicular.
7. ( ) Angle of incidence at which light may not pass into air from a denser medium.

(2) Part II—Knowledge of a Principle

In optics the angle of reflection is:

a. equal to the angle of refraction.
b. less than the critical angle.
c. equal to the angle of incidence.
d. equal to the angle of minimum deviation in the case of a prism.
e. a minimum for light incident on a plane surface.

(3) Part III—Application of a Principle

Highway engineers wish to know the average velocity of cars traveling along a certain highway. A mirror is placed so that when a car crosses a line, its image is reflected to an observer stationed on the road at a known distance from the line. By means of a stop watch he measures the time the car takes to travel from the first line mentioned to a line across the road directly in front of him. How should the mirror be placed?

a. The mirror should be perpendicular to the observer.
b. The mirror should be perpendicular to the line across the road.
c. The mirror should be at an angle for forty-five degrees to the observer.
d. The line drawn perpendicular to the mirror should bisect the angle made by a line to the observer and the line across the road.
e. The mirror should be placed half way between the observer and the line across the road.

It is noted that the vocabulary test (1) calls for a knowledge of the technical terms employed in Part II and Part III; the test item (2) calls for a knowledge of the physical principle that the angle of incidence is equal to the angle of reflection; and the test item (3) calls for the application of this principle.

An illustration drawn from an examination in Human Biology[2] testing for the knowledge of a principle (2) and its application (3) is:

(2) Part II—Knowledge of a Principle

The principle of bone construction is essentially:

a. the maintenance of symmetry.
b. the ability to withstand sudden shocks.
c. the best distribution of physiological stress.
d. the largest marrow cavity possible for blood formation.
e. the maintenance of large lime deposits.

(3) Part III—Application of a Principle

At the World's Fair, an artificial skeleton was exhibited which performed all of the motions of the body, including hitting, jumping, and running. The great construction principle of the artificial bones was:

a. to duplicate the external contour of natural bones.
b. to duplicate the chemical composition of natural bones.
c. to duplicate the diameter and length of the marrow cavities of normal bones.
d. to duplicate natural lines of stress and strain.
e. the ability to withstand sudden shocks.

A third illustration from an examination in Basic Wealth is:

(2) Part II—Knowledge of Facts

Of the places mentioned below the one that would be relatively warmest on a still evening after a hot summer's day is

a. a hay field in a relatively flat region.
b. the down-town section of a large city.
c. a suburb of a large city.
d. a farm on the south slope of a hill.
e. a typical bituminous coal mine.

(3) Part III—Application

Observers for the United States Weather Bureau report that in summer the late evening temperatures in suburbs are often from ten to twenty degrees lower than the temperatures recorded

[2] The vocabulary tests for the illustrations in Human Biology and Basic Wealth are not presented here. In these instances, a unit of vocabulary consisted of the technical terms in several test items of Parts II and III.

at the same time in the industrial section of the adjacent city. The explanation of this is most closely connected with the following fact:

a. Temperature variations are greatly influenced by altitude differences.
b. Suburbs are closer to lakes or other large bodies of water.
c. Wind is more common in the suburbs.
d. There are more trees in the suburbs.
e. Artificial structures give off heat by radiation for a considerable time after sunset.

It should be pointed out that the three examinations were separately assembled and that Parts I and II were given in the morning and Part III was given in the afternoon of the same day. In this way, questions in one examination did not provide suggestions for the answer to questions occurring in another. Each of the examination periods was three hours in length.

THE ANALYSIS OF THE MEASURES

From the description of the examinations and their method of construction attention is now turned to the analysis of the test results. Although similar data are available for courses in Astronomy and Technology, the findings presented here will be limited to those resulting from the examinations in three courses: Physics and Chemistry, Human Biology, and Basic Wealth. The coefficients of reliability of the whole examinations in these fields were: Physics and Chemistry, $r = .96 \pm .006$; Human Biology, $r = .95 \pm .006$; and in Basic Wealth, $r = .98 \pm .003$ (See Table I).

INTERRELATIONS OF MEASURES

The nature of the examination program made possible some exploration into the problem of the interrelations of the measures on the several parts of the examinations. Since we are interested in the intrinsic relation or association of the abilities measured, the raw coefficients of the correlation have been corrected for attenuation.

The examinations in both Physics and Chemistry and Basic Wealth consisted of three parts. Part I was made up of vocabulary items; Part II of items dealing with principles and other factual information; and Part III, of items calling for interpretation and application of principles. The coefficients of correlation between the measures of the acquisition of vocabulary and of principles and other factual information were: $.82 \pm .03$ (atten. $r = .85 \pm .02$) in Physics and Chemistry, and

TABLE I

RELIABILITY COEFFICIENTS OF THE EXAMINATIONS

Examination of	Chemistry and Physics	Sp. Br. r_{11} Basic Wealth	Human Biology
Part I Terminology	.96 ± .006	.96 ± .007	.96 ± .006
Part II Principles and other Factual Information	.91 ± .015	.94 ± .01	
Part III Interpretation	.86 ± .012	.83 ± .03	.64 ± .04
Whole Examination	.96 ± .006	.98 ± .003	.95 ± .006
Part II View points on Health and Disease			.80 ± .03
Number of Cases	77	54	74

TABLE II

INTERCORRELATIONS OF THE SEVERAL PARTS OF THE COMPREHENSIVE EXAMINATION

Examination of	Part II Principles and Other Factual Information r	P.E.	Atten. r	P.E.	Part III Interpretation and Application r	P.E.	Atten. r	P.E.	No. of Cases
Part I—Terminology									
Chemistry and Physics	.82	± .03	.85	± .02	.75	± .03	.82	± .03	79
Basic Wealth	.70	± .04	.74	± .04	.67	± .05	.77	± .04	54
Human Biology	.....		.....		.....		.....		..
Part II—Principles and other Factual Information									
Chemistry and Physics	.....		.....		.77	± .03	.87	± .02	78
Basic Wealth	.....		.....		.78	± .04	.89	± .02	54
Human Biology	.....		.....		.60	± .05	.77	± .03	73
Part II—Health and Disease Situations									
Human Biology	.29	± .07	.34	± .06	.41	± .06	.58	± .05	72

.70 ± .04 (atten. r = .74 ± .04) in Basic Wealth. Other intercorrelations were: between vocabulary and interpretation or application .75 ± .03 (atten. r = .82 ± .03) in Physics and Chemistry; .67 ± .05 (atten. r = .77 ± .04) in Basic Wealth; between the measures of the acquisition of principles and the measures of interpretation and application: .77 ± .03 (atten. r = .87 ± .02) in Physics and Chemistry; .78 ± .04 (atten. r = .89 ± .02) in Basic Wealth (Table II).

The organization of the examination in Human Biology differed somewhat from that of the examinations for the two courses described above. Part I consisted of test items dealing with principles and other factual information including vocabulary. Part II was made up of items dealing with various aspects of health and disease; Part III was comprised of items calling for interpretation and application of principles. The coefficients of correlation between the several parts were: Principles and other factual information, and interpretation and application, .60 ± .05

(atten. r = .77 ± .03); health and disease situations, and principles and other factual information, .29 ± .07 (atten. r = .34 ± .06); health and disease situations, and interpretation and application, .41 ± .06 (atten. r = .58 ± .05).

The magnitude of most of the coefficients of correlation between the various measures point to the inference that the ability to acquire information on the part of students in these courses has been accompanied to a substantial extent by the ability to apply this information.

THE SEPARABLE AND COMBINED IMPORTANCE OF CERTAIN VARIABLES

A further method of analysis makes possible the measure of the combined importance of the several independent variables in estimating the dependent variable. If we consider (1) the measure of the ability to interpret and apply principles and other factual information as the dependent variable, and (2) the measure of vocabulary, (3) the measure of the acquisition of principles and other factual information, and (4) the scores on the Minnesota College Aptitude Test, as dependent variables, the multiple correlations are:

For Physics and Chemistry—
$R_{1.234} = .803$ Atten. $R_{1.234} = .87$
For Basic Wealth—
$R_{1.234} = .831$ Atten. $R_{1.234} = .94$
For Human Biology—
$R_{1.34} = .749$ Atten. $R = .91$

The number of cases entering into these computations were those students for whom all measures were available: 74 in Physics and Chemistry, 48 in Human Biology, and 72 in Human Biology.

When the multiple correlations are squared, an index is obtained of the proportion of the variance in the dependent variable which is mathematically accounted for. In the instance of the above multiple correlations this proportion becomes 76 percent in Physics and Chemistry, 88 percent in Basic Wealth, and 83 percent in Human Biology.

A measure of the separable importance of each of the independent variables is obtainable by means of the multiple regression equation where the standard deviation of both the independent variables and the dependent variables is taken as unity. These equations for our problems are:

Physics and Chemistry—

$$X_1 = .3730\,x_2 + .6211\,x_3 + .1755\,x_4 + K$$

Basic Wealth—

$$X_1 = .2172\,x_2 + .6433\,x_3 + .2553\,x_4 + K$$

Human Biology—

$$X_1 = .7077\,x_2 + .4700\,x_4 + K$$

The percentage of influence (or association) of each of the independent variables upon (or with) the variance of the dependent variable is given by the squares of the regression coefficients. The values and the percentages or weight for our problems are:

Physics and Chemistry—	*Variable Regression Coefficient Squared*	*Percentage or Weight*
2. Vocabulary	.1391	25.0
3. Principles, etc.	.3858	69.4
4. College Aptitude	.0308	5.5
Total	.5557	99.9

Basic Wealth—	*Variable Regression Coefficient Squared*	*Percentage or Weight*
2. Vocabulary	.0471	8.9
3. Principles, etc.	.4138	78.6
4. College Aptitude	.0652	12.4
Total	.5261	99.9

Human Biology—	*Variable Regression Coefficient Squared*	*Percentage or Weight*
2. Vocabulary and 3. Principles, etc.	.5008	69.4
4. College Aptitude	.2209	30.6
Total	.7217	100.0

So far as the factors considered contribute toward the ability to apply knowledge, knowledge of vocabulary, knowledge of principles and other factual information, and the ability measured by the college aptitude test contribute with the relative weight of 4.5 : 12.6 : 1 in Physics and Chemistry and of 1 : 8.8 : 1.4 in Basic Wealth. Knowledge of principles, vocabu-

lary, and other factual information (combined) and the ability measured by the college aptitude test contribute with the relative weight of 2.3 : 1 in Human Biology.

The squares of the regression coefficients give the measures of the single importance of the independent variables. Since these variables are correlated some measure of their joint contributions is desirable. An estimate of this joint contribution is attainable through the application of the method of path coefficients developed by Sewell Wright.[3] Through the application of this technique, the net joint contribution of a knowledge of vocabulary, a knowledge of principles and other factual information, and college aptitude as measured by the college aptitude test was found to be: .2309 in Physics and Chemistry and .3141 in Basic Wealth.

It may be well to point out again to the reader that this analysis has been limited to a consideration of three of the measurable outcomes of instruction. Obviously there are other outcomes which have not been considered and for which highly reliable and valid measures are not yet available, *i. e.*, the development of interest in science, the ability to use the scientific method, the development of scientific attitudes, the ability to manipulate laboratory apparatus, or laboratory resourcefulness, increased powers of observation, etc. The outcomes to be given most importance in examinations will depend upon those held of highest purpose in instruction. For instance, in this study if the ability to apply a knowledge of vocabulary, principles, or other factual information to the solution of problems or situations be considered the end-product or the sole purpose of instruction, then it might be argued that it alone needs to be tested for. However, if a measure of growth in scientific vocabulary or the acquisition of important principles is wanted, separate measures would be necessary for their determination. Even if it be accepted that functional knowledge is the end-product, it will likely be of considerable value for diagnostic and instructional purposes to know the competency of the student with respect to other abilities and the relation they bear to one another and to the end-product itself. Likewise in the determination of the relative weightings by means of the multiple regression equations, such additional factors as initial status of students, maturity, reading abilities, sex, and others might have been considered, so that it should be kept in mind here as well that the weightings are with reference to the independent variables introduced.

[3] Wright, Sewell. "Correlation and Causation," *Journal of Agricultural Research*, XX (1921), pp. 238-255.

SUMMARY

We may now sum up the main points of this discussion of the measurement of differential functions.

1. Valid examinations should aim to measure all the purported outcomes of instruction that are amenable to measurement.

2. The outcomes of instruction should be defined in specific terms of what the student is expected to do to reveal the extent of his acquisition.

3. The process of developing examinations directed at the measurement of three outcomes of instruction has been described.

4. While the interrelations of the abilities tested were high in some cases and substantial in others, separate measures appear essential in order to make a more accurate appraisal of student accomplishment.

5. The weighting in order of the separable importance in determining the ability to apply knowledge is (1) knowledge of informational facts and principles, (2) vocabulary, and (3) the measure of college aptitude used.

RELATION BETWEEN INTELLIGENCE AND CERTAIN ASPECTS OF PHYSICAL ABILITY

J. C. Seegers and Otto Postpichal
Temple University

This study considers the relationship between intelligence and certain phases of physical achievement. It was undertaken because of the feeling apparently prevalent among many Philadelphia teachers of physical education that they found lower levels of achievement in physical activities in the schools organized primarily to care for children of lower mentality.

Previous studies seemed inadequate in one or more or several particulars. In some instances the definition of the population of the studies was not precise; in others, the measure of physical ability was derived from scores achieved in several events, necessitating the establishment of an index number, derived by weighting the several events. No such index numbers were presented. Another defect was that averages and means without measures of dispersion were presented.[1]

THE GENERAL METHOD OF THE STUDY

The subjects of the study were 656 boys in two special schools in the city of Philadelphia. Only boys were considered in order to eliminate sex as a variable. Because of the obvious effect of serious degrees of over weight or under weight conditions, all individuals thirteen percent under weight or over weight, according to the Baldwin-Wood Tables, were excluded. A mental score for each boy under consideration was obtained from the offices of Philadelphia schools. Each boy was given athletic tests by the physical education teacher in the following events: standing broad jump; overhead ball throw; 50-yard dash; standing hop, step, and jump; chinning. The better result of two trials was recorded as the result or score except in the dash and in chinning, in which only one trial was given. These tests were repeated twice, after intervals of one month.

The reliability of these athletic tests was checked by computing correlations with two sets of scores for each of the events. The coefficients ranged from .82 to .92, indicating a high degree of reliability.[2]

[1] Paterson, Donald G. *Physique and Intellect.* (New York: Century Co., 1930), especially page 169 and pages 284-86.

[2] Kelley, Truman Lee. *Interpretation of Educational Measurements.* New York: World Book Co., 1927. According to Kelley, the minimal satisfactory reliability, as determined by a reliability coefficient, of a test for measuring a school group in some specific subject and estimating future group promise in the same or a closely related subject is .5.

In order that the influence of age might be reduced to a minimum the subjects were grouped chronologically at step intervals of one year. For example all of the boys over 13 years 11 months of age and not over 14 years 11 months of age were grouped together. The relationship between intelligence, as indicated by I.Q.'s, and athletic ability, as measured by each of the five events, was calculated separately for each age group.

These calculations were computed on the basis both of average results scored in each event, and of best results achieved during the three months. It is of interest to note that the distributions obtained by the two methods did not differ materially.

The calculations included correlations between I.Q.'s and scores in each event on each age level and determination of the forecasting efficiency indicated by each coefficient of correlation. All correlations were computed by the Pearson product-moment method, and in each case Blakeman's test[3] indicated the relationship was linear.

PRESENTATION OF DATA

In Table I the mean and standard deviation of the I.Q.'s of each age group are given; in Table II, the means and standard deviations of records of the events.

Particular attention should be given to the sigma deviations in each of these tables. They indicate definitely the dangers of comparisons based only upon means or averages. It may be noted for instance in Table II that the mean of any age group does not differ by more than one sigma from the mean of the preceding or succeeding group. The means and standard deviations exhibited in Table I show no significant difference in the mental ability of the age groups.

TABLE I

MEANS AND STANDARD DEVIATIONS OF THE DISTRIBUTIONS OF I.Q.'S OF 656 BOYS IN THE INDICATED AGE GROUPS

Age Group	Number of Cases	Mean	Sigma
16	81	72.7	9.65
15	135	73.3	9.45
14	135	73.4	9.35
13	125	73.6	9.10
12	100	75.3	9.30
11	80	76.2	9.75

[3] Garrett, Henry E. *Statistics in Psychology and Education.* (New York: Longmans, Green & Co., 1926), page 210.

TABLE II

MEANS AND STANDARD DEVIATIONS OF RECORDS OF THE EVENTS FOR THE INDICATED AGE GROUPS

EVENTS Age.....	11	12	13	14	15	16
Jump			Inches			
Average Mean	63.7	66.6	70.3	73.9	78.4	80.4
Score Sigma	6.7	6.2	7.5	7.8	8.5	7.9
Best Mean	65.6	69.0	71.8	76.2	80.4	83.5
Score Sigma	6.4	6.5	7.5	8.2	8.8	8.0
Throw			Feet			
Average Mean	24.1	26.0	28.9	31.5	34.5	36.9
Score Sigma	3.8	3.7	4.4	4.9	5.3	4.8
Best Mean	25.5	27.4	30.5	33.0	35.2	38.6
Score Sigma	3.4	3.9	4.7	4.9	5.3	5.0
Dash			Seconds			
Average Mean	8.16	7.98	7.71	7.48	7.33	6.92
Score Sigma	.50	.49	.52	.55	.56	.51
Best Mean	8.00	7.78	7.57	7.31	7.03	6.80
Score Sigma	.46	.44	.53	.51	.58	.51
Hop-Step-Jump			Feet			
Average Mean	14.2	15.3	15.6	17.1	18.2	18.7
Score Sigma	1.9	2.2	2.0	2.5	2.6	2.4
Best Mean	15.2	16.1	16.5	17.9	19.3	19.6
Score Sigma	1.9	2.1	2.1	2.4	2.5	2.6
Chinning			Times			
Average Mean	4.2	4.8	5.1	5.6	7.0	8.1
Score Sigma	3.0	2.9	2.8	3.2	3.4	2.6
Best Mean	5.2	5.6	5.8	6.5	7.9	8.7
Score Sigma	2.5	3.2	3.1	3.4	3.5	2.8

Tables III[4] and IV, displaying the performance standards set respectively for general and for special classes in Philadelphia schools, are provided for purposes of comparison. It will be noted that the average achievement indicated in Table II in each instance exceeds that expected of special classes, and closely approximates that demanded for a grade

TABLE III

PHILADELPHIA AGE AIMS (BOYS) IN PHYSICAL EDUCATION FOR NORMAL CLASSES

Age	Standing Broad Jump Feet Inches			Dodgeball Far Throw Feet			50 Yard Dash Seconds Fifths			Chinning Times		
Years	F	G	E	F	G	E	F	G	E	F	G	E
16	5-10	7-1	7-10	31	39	45	7-4	6-4	6-2	2	7	8
15	5-8	6-9	7-6	28	36	42	8-0	7-0	6-3	2	6	8
14	5-5	6-5	7-2	25	33	39	8-1	7-1	6-4	2	5	7
13	5-1	6-1	6-10	23	30	35	8-2	7-3	7-0	1	5	6
12	4-11	5-10	6-6	22	28	32	8-3	7-4	7-1	1	4	6
11	4-7	5-7	6-2	20	26	30	8-4	8-9	7-2		3	5
10	4-5	5-3	5-10	19	24	28	9-0	8-1	7-3		3	5
9	4-2	5-0	5-7	17	22	26	9-1	8-2	7-4		2	4
8	3-10	4-9	5-4	16	20	24	9-2	8-3	8-0		1	3

[4] The basis for the grades is said to be that experience and observation indicate the upper 10 percent should reach E standards, the next 40 percent G standard, the next 40 percent F, the remaining 10 percent receiving a grade of unsatisfactory.

of G in general classes. Grade F achievement is exceeded consistently by the scores of Table II.

TABLE IV

PHILADELPHIA AGE AIMS (BOYS) IN PHYSICAL EDUCATION FOR SPECIAL CLASSES

Age	Standing Broad Jump Feet	Standing Broad Jump Inches	Ball Throw Overhead Feet	50 Yard Dash Seconds	50 Yard Dash Fifths	Chinning Times	Hop-Step-Jump Feet
16	5	8	28	8	0	3	15
15	5	6	28	8	2	2	14
14	5	2	24	8	3	2	13
13	5	0	23	9	0	2	12
12	4	6	21	9	1	2	12
11	4	4	20	9	3	1	11
10	4	3	18	9	4	1	11
9	4	0	16	9	4	1	9
8	3	10	13	10	1	1	7

In Table V correlations of scores and I.Q.'s within each age group are displayed for each event, as well as the coefficient of alienation and its complementary index for forecasting, which indicates that in no case is the predictive value high. There is no justification with correlations of the order obtained for assuming that the measures are sufficiently related to permit the establishment of standards within groups of comparable chronological age. It is significant, however, to note that the highest correlations were secured between I.Q. and the hop-step-jump event; the lowest between I.Q. and chinning. It seems that the more complicated events, *i. e.*, "complicated" in terms of muscle groups involved and coördination demanded, are correlated more highly with intelligence as defined by these I.Q.'s than are less complicated events.

An extension of the study was conducted with the data from the 70–74 I.Q. group, selected because its population (135 cases) was the greatest found in any five point I.Q. step interval. For this group the correlation between chronological age and scores in the five events, with the corresponding values of K, the coefficient of alienation, and the consequent forecasting index were computed. The results are displayed in Table VI. In Table VII in which these results are compared with those based upon I.Q. score correlations, it will be noted that the predictive value of age is much greater.

TABLE V

COEFFICIENTS OF CORRELATION AND ALIENATION WITH FORECASTING INDICES FOR THE RELATIONSHIP BETWEEN I.Q.'S AND SCORES IN ATHLETIC EVENTS FOR THE INDICATED AGE GROUPS

EVENTS — Age	11	12	13	14	15	16
Average Jump						
Coeff. of Correlation	.35	.44	.49	.46	.44	.49
Probable Error	.07	.05	.05	.05	.05	.05
Coeff. of Alienation	.94	.89	.87	.89	.90	.87
Forecasting Index	6	11	13	11	10	13
Best Jump						
Coeff. of Correlation	.30	.47	.49	.47	.40	.49
Probable Error	.07	.05	.05	.05	.05	.06
Coeff. or Alienation	.95	.88	.87	.88	.92	.87
Forecasting Index	5	12	13	12	8	13
Average Throw						
Coeff. of Correlation	.31	.38	.41	.43	.38	.47
Probable Error	.07	.06	.05	.05	.05	.06
Coeff. of Alienation	.95	.93	.91	.90	.93	.88
Forecasting Index	5	7	9	10	7	12
Best Throw						
Coeff. of Correlation	.35	.34	.42	.43	.42	.41
Probable Error	.07	.06	.05	.05	.05	.06
Coeff. of Alienation	.94	.94	.91	.90	.91	.91
Forecasting Index	6	6	9	10	9	9
Average Dash						
Coeff. of Correlation	.30	.32	.38	.37	.35	.34
Probable Error	.07	.06	.05	.05	.05	.07
Coeff. of Alienation	.95	.95	.93	.93	.94	.94
Forecasting Index	5	5	7	7	6	6
Best Dash						
Coeff. of Correlation	.33	.32	.35	.36	.34	.35
Probable Error	.07	.06	.05	.05	.05	.07
Coeff. of Alienation	.94	.95	.94	.93	.94	.94
Forecasting Index	6	5	6	7	6	6
Average Hop-Step-Jump						
Coeff. of Correlation	.45	.50	.50	.45	.52	.55
Probable Error	.06	.05	.05	.05	.04	.05
Coeff. of Alienation	.89	.87	.87	.89	.85	.84
Forecasting Index	11	13	13	11	15	16
Best Hop-Step-Jump						
Coeff. of Correlation	.45	.50	.50	.45	.52	.55
Probable Error	.06	.05	.05	.05	.04	.05
Coeff. of Alienation	.89	.87	.87	.89	.85	.84
Forecasting Index	11	13	13	11	15	16
Average Chinning						
Coeff. of Correlation	.19	.15	.20	.27	.20	.24
Probable Error	.07	.07	.06	.05	.06	.07
Coeff. of Alienation	.98	.99	.98	.96	.98	.97
Forecasting Index	2	1	2	4	2	3
Best Chinning						
Coeff. of Correlation	.23	.14	.21	.25	.15	.22
Probable Error	.07	.07	.06	.05	.06	.07
Coeff. of Alienation	.97	.99	.98	.97	.99	.98
Forecasting Index	3	1	2	3	1	2

TABLE VI

COEFFICIENTS OF CORRELATION BETWEEN CHRONOLOGICAL AGE AND EVENT SCORES AMONG 135 CHILDREN OF I.Q. 70–74

EVENT	Coeff. of Correlation	Probable Error	Coeff. of Alienation	Forecasting Index
Jump	.71	.03	.70	30%
Throw	.70	.03	.71	29%
Dash	.73	.03	.68	32%
Hop-Step-Jump	.64	.03	.77	23%
Chinning	.35	.05	.94	6%

TABLE VII

COMPARISON OF FORECASTING INDICES DERIVED FROM THE RELATIONSHIP BETWEEN I.Q.'S AND CHRONOLOGICAL AGE WITH EVENT SCORES MADE BY 135 CHILDREN OF I.Q. 70–74

EVENT	Percent of Forecasting Efficiency of I. Q.	Percent of Forecasting Efficiency of Chronological Age
Jump	11	30
Throw	8	29
Dash	6	32
Hop-Step-Jump	13	23
Chinning	2	6

SUMMARY AND CONCLUSIONS

The study is one of correlation between I.Q.'s and scores in each of five athletic events achieved by 656 boys in the public schools of Philadelphia. The correlations between I.Q.'s and scores in the athletic events were positive but too low to be of much forecasting usefulness. The I.Q. score correlations were higher for the more complicated events. These facts indicate a definite correlation between the presumably desirable attributes of intelligence as measured in the study and the athletic ability tested in the five events. Brighter boys tend to achieve better scores. However, the individual variation is so great, and the forecasting utility is so small that these tendencies should not be given an individual application.

(NOTE: Readers interested in the exact statistical procedure, including demonstration of the linearity of the correlation, significance of the coefficients, and similar details omitted in the interest of space are referred to: Postpichal, Otto. *The Relation between Intelligence and Athletic Ability in Cases of Orthogenic Backward Boys.* Typewritten Master's Thesis, Temple University, Philadelphia.)

IMPROVEMENT IN EDUCATIONAL ACHIEVEMENT ACCOMPANYING A HEALTH EDUCATION PROGRAM*

MARTHA CRUMPTON HARDY
Elizabeth McCormick Memorial Fund

AFTER two years' participation in a general health education program conducted in Grades I and II, pupils in Grade III (both high and low sections) were selected for an experimental investigation on the measurable effects of health teaching projects. Pupils from twelve schools of the public school system of a small industrial city were included in the experimental study. In eight of these schools an extensive three-year health education program was carried on, which consisted of special classroom instruction on the principles of healthful living, conferences with parents and teachers by a staff of child specialists, individual pupil interviews, periodic physical examinations by a pediatrician in the presence of the parent, and follow-up work by the school nurses. Children in these schools are herein referred to as the health instructed or experimental group. In the remaining four schools the pupils were given yearly physical examinations but had no special health teaching other than the prescribed course in hygiene. These pupils formed a control group against which the improvement of the experimental group could be compared.

Pupils were selected for the study primarily on the willingness of the parents to coöperate in a long time program. At the conclusion of the investigation the records showed 409 of the 452 originally selected pupils with complete physical, psychological, and educational data. Of these, 268 had remained in the health instructed classes, 104 had continued throughout in the non-instructed classes, and 37 had transferred from one group to the other.

Educational achievement was measured by the Stanford Achievement Tests which were administered annually in May. Over 800 pupils took the test for three consecutive years. At the time of the first educational test 69 percent of the selected cases were in Grade III and 31 percent had advanced to Grade IV-B. At the third test 68 percent were in Grade V, 27 percent had advanced to Grade VI, and 5 percent were in Grade IV.

The school situation for the groups under observation was as uniform as possible and from the educators' point of view there was no reason

* These data are from an investigation conducted in the public schools of Joliet, Illinois. The general report of the study will appear in the fall of 1936 under the title *Healthy Growth*, published by the University of Chicago Press.

to expect any variation in school accomplishment. Detailed yearly reports of the test results from each class examined were made to the superintendent of the elementary schools. It would, therefore, seem quite unlikely that the teachers in the control schools were any less interested in having good test records for their pupils than were the teachers of the experimental group. Any special subject testing, other than that of the investigation, was planned by the school supervisors and affected all schools equally. Some teachers of the classes given health instruction were greatly interested in the program and, unquestionably, gave a strong incentive to their pupils. But the teachers in the control appeared to be as much interested in their work as the other teachers, and did as good or better general teaching, which after all is a very significant factor in pupil achievement.

Reports and recommendations relative to pupil abilities and disabilities were made by the psychologist to the principals and supervisors of all the pupils examined by her, including discards from the major study as well as the cases forming the control and experimental groups. The control thus had the benefit of these special suggestions, although they comprised the only efforts made to include aspects of mental hygiene in the health teaching program. In fairness to the children and as a means of insuring continual coöperation of the teaching staff of the control schools, this concession was deemed expedient. It was, of course, a psychological service which should follow any efficient testing program and has been treated as such, but it gives further evidence of the uniformity in the school situations of the groups under observation.

From the test results it appeared likely that the selected cases were brighter and more successful in school than the pupils who were not included in the study. They were also somewhat younger. Table I pre-

TABLE I

EDUCATIONAL STATUS OF THE SELECTED AND DISCARDED CASES AT THE INITIATION OF THE HEALTH PROGRAM

Groups	N	Age M	Intelligence Quotient M	Grade Av.	Educational Quotient M	Reading Quotient M	Arithmetic Quotient M	Spelling Quotient M
Selected Cases								
Health instructed	268	9-2	105	3A	109	108	108	111
Transfers	37	9-4	99	3A	106	104	104	108
Control	104	9-5	96	3A	105	103	106	109
Discards	416	9-9	97*	3A	100	97	101	102

*209 of these cases had the Stanford Revision Intelligence Test which was given annually to the selected cases.

sents the general educational status of 825 pupils in the twelve elementary schools at the time of the initial tests. On the average, the pupils to whom special health instruction was given had the highest scores.

Since the groups were not equally successful when the projects were begun, it was necessary to make them so before the influence of the health work on school improvement could be determined. For comparative purposes, the two main groups under observation were equated with respect to initial test score and age. The procedure followed was that of making the distributions proportionate, which resulted in eliminating the low scoring and older pupils from the control group and the high scoring and younger pupils from the experimental group. The elimination was done by drawing the required number from each age, sex, and score distribution, the ratio between the groups being kept constant. The names were written on slips, which were placed in boxes classified according to pupil sex, age level (12 month interval) and initial test score (10 point interval) and the drawing done by pure chance. The results are presented in terms of the total educational score and of the scores for the special subjects of reading, arithmetic, and spelling expressed as ages or increments of improvement. The reported data are limited to these three subjects because the content of the primary examination of the test is comparable to the advanced form in only these subjects.

RATE OF EDUCATIONAL DEVELOPMENT DURING THE HEALTH PROGRAM

Development curves have been plotted from the educational age data and are shown in Figure 1. The superiority of the health instructed group to a control of like ages and initial accomplishment is strikingly revealed in these curves. It can be stated with a high degree of confidence that the pupils who participated in the health projects showed greater improvement in their school subjects during this period than a control group. The critical ratio of the difference corrected for correlation of retests was 3.3 in standard error units.

For most of the cases, the interval between tests was an exact calendar year, but, of course, there were some irregularities. In order to eliminate the possibility that group inequalities were due to variations in the testing procedure, an average monthly increment of achievement was determined for each pupil. This was done by dividing the total gain, in terms of educational age, by the length of time intervening between the dates of testing. For illustration, if a pupil's score on the first test was 110 educational months and that on the last, 138 months, his gain would be 28

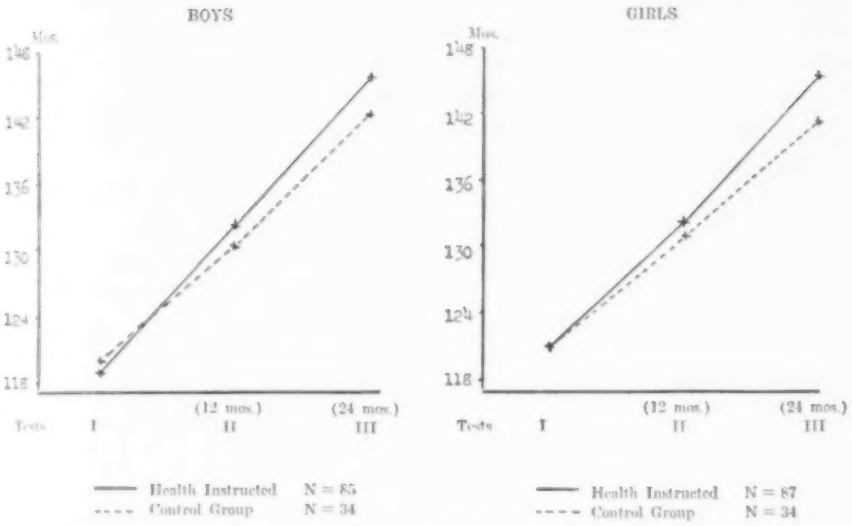

FIGURE 1: DEVELOPMENTAL CURVES MEASURED IN COMPOSITE EDUCATIONAL AGE, SHOWING SCHOOL IMPROVEMENT DURING A HEALTH PROGRAM WHEN THE GROUPS WERE EQUALLY SUCCESSFUL ON THE FIRST TEST AND OF LIKE AGES

months and, if the time interval was 25 months, then his average growth increment would be $\frac{28}{25}$ or 1.12 educational months to each calendar month.

An inspection of the results expressed in terms of this monthly increment of growth showed conclusively that the health instructed group had the higher rate of accomplishment during the years covered by this investigation, the critical ratio of the group difference being greater than 3.00. With respect to particular school subjects, the differences were less marked than those for general achievement but their trend was very uniform.

From the age comparisons, it was apparent that, age for age, the accomplishment rates were higher for the health trained group. Out of thirty-two age and sex comparisons, the rate of the health trained was exceeded or equaled by that of the control group in only seven instances, five of these occurring with the eleven year old group, which had a very small number of cases, and the other two with respect to the arithmetic scores of the nine year olds.

These results on school accomplishment were from groups made comparable with respect to age, sex, and initial educational score. Ratings

on the intelligence tests indicated that they were also equally bright. A careful inspection of the yearly I.Q.'s, however, established a difference between the groups, the health trained having the higher scores (critical ratio, 3.02 standard error units). Consequently, the instructed group, which was composed of the more capable pupils should have made more rapid progress.

This difficulty was met by equating the groups with respect to the four variables of age, sex, I.Q., and initial achievement. While the size of the groups was reduced by this procedure, the distributions remained proportionally the same, the experimental group containing 2.5 times as many pupils as the control. In every case, the size of the control was greater than 50.

The ability classifications were based first, on the initial I.Q. and the initial achievement score, when this was judged in terms of the composite educational age. These findings were then checked by using the average of three I.Q.'s, as a more satisfactory estimate of general intelligence, and the initial scores in the school subjects, as the basis for comparing progress during the investigation. The similarities between the groups under these conditions were without question, *e. g.*, the average child in each group had a rating on the first tests of 100 as his I. Q. and 9 years 10 months, as his educational maturity. There were differences, however, with regard to the achievement in the various subjects. When the initial accomplishment in these was taken into consideration in the equating procedures, the variations in no case were as large as two months and wherever they occurred, the advantage was uniformly given to the control group.

In the findings from these comparisons, the average gain of the health instructed group, without exception, was larger in each of the school subjects than that of their control, when the factors of ability and initial achievement were both held constant. It would seem very probable that a true difference was indicated in regard to general achievement, since the critical ratio was as large as 2.36, when the rate of improvement was uncorrected for correlation of retests. Certainly the probabilities would be higher than 99 to 1 for the more rapid progress of the instructed pupils. Complete reliability was established in so far as differences in spelling proficiency were concerned. (See Table II.)

A certain amount of overlapping in the school achievement of different grades is commonly expected. Consequently, when two groups are alike with respect to age and achievement, it is not always the case that

TABLE II

GROUP AND AGE COMPARISONS IN TERMS OF THE AVERAGE MONTHLY INCREMENT OF SCHOOL ACHIEVEMENT WHEN AGE AND ABILITY WERE CONTROLLED

	Age and Initial Achievement Constant									Age, I.Q., and Initial School Score Constant								
		Educational Increment		Reading Increment		Arithmetic Increment		Spelling Increment			Educational Increment		Reading Increment		Arithmetic Increment		Spelling Increment	
	N	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	N*	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D
8 YEARS																		
Instructed	38	1.28	.27	1.58	.44	1.37	.46	1.10	.34	26	1.25	.27	1.45	.40	1.18	.26	1.13	.33
Control	15	1.01	.17	1.32	.43	1.15	.35	.88	.30	10	.93	.10	1.23	.36	.97	.34	.90	.28
9 YEARS																		
Instructed	81	1.09	.31	1.29	.47	1.29	.43	.98	.27	96	1.08	.30	1.28	.44	1.24	.50	.98	.29
Control	31	1.02	.26	1.17	.45	1.33	.40	.91	.23	37	1.06	.24	1.25	.48	1.34	.38	.86	.29
10 YEARS																		
Instructed	44	.92	.31	1.00	.42	1.13	.46	.92	.32	43	.88	.31	.95	.42	1.12	.40	.89	.31
Control	18	.78	.21	.79	.40	1.06	.52	.63	.22	17	.76	.24	.73	.34	.97	.39	.68	.18
11 YEARS																		
Instructed	9	.86	.34	.77	.61	1.32	.64	.81	.26	8	.78	.37	.55	.37	1.30	.75	.81	.41
Control	4	.71	.08	.96	.29	.84	.17	.57	.16	4	.81	.21	.86	.20	1.21	.23	.77	.25
BOYS																		
Instructed	85	1.10	.35	1.29	.54	1.26	.49	.98	.34	84	1.07	.35	1.19	.50	1.19	.45	.96	.33
Control	34	.97	.25	1.17	.45	1.16	.42	.79	.31	34	1.00	.24	1.15	.46	1.17	.43	.83	.29
GIRLS																		
Instructed	87	1.05	.31	1.22	.48	1.28	.44	.99	.27	89	1.02	.30	1.19	.47	1.22	.48	.99	.30
Control	34	.90	.25	1.01	.47	1.22	.45	.82	.25	34	.90	.26	1.04	.49	1.21	.38	.81	.25
BOYS AND GIRLS																		
Instructed	172	1.07	.34	1.25	.52	1.27	.46	.98	.31	173	1.04	.33	1.19	.48	1.21	.46	.98	.31
Control	68	.93	.26	1.09	.47	1.19	.44	.81	.28	68	.95	.26	1.10	.47	1.19	.41	.82	.27
Critical Ratio**		3.50		2.35		1.26		4.25			2.36		1.34		.33		4.00	

*Number of cases varied slightly for the different subject comparisons.
**Uncorrected for correlation of re-test.

they are in the same grades. If, in equating the groups, a larger proportion of the instructed than of the control had been selected from the highest grade level, then extra drill and greater familiarity with the subject matter might have accounted for their more rapid progress during the period of this study. That this was not the situation, was found when the age-grade records were examined.

Age for age, the control group were more frequently from the advanced grade and, where age variations occurred, the control were the younger. While the differences were not large, fewer among the control were in the lowest grade (3B) and more were in the highest grade (4B) at the time of the first test, than among the instructed group, although the two groups were, on the average, of the same age, intelligence, and achievement. Inspection of the age-grade distributions also showed that a greater proportion of the youngest pupils from the control than from the instructed were doing advanced work.

The comparative results have clearly indicated that, when a group given health training were of the same age and intelligence as a control group, although somewhat less advanced in school and having a lower achievement record, the health instructed pupils tended to show greater improvement in their regular school work. It is quite possible that the few exceptions to this general trend were outcomes of the minor handicaps placed upon the experimental group, which a short-time program could not always overcome.

IMPROVEMENT UNDER CONDITIONS OF THE INVESTIGATION COMPARED WITH SCHOOL STANDARDS

As the school tests were given to all pupils in the classes selected for study, irrespective of their connection with the experiment, it was possible to develop local norms evaluated for the practice factor involved in the re-examinations. Hence the improvement of the pupils under observation can be contrasted with the expected achievement of pupils at their grade level and with the actual achievement of their classmates.

According to the grade standards for the New Stanford Achievement norms, a pupil progressing from 3A through 5A should have gained 24 to 25 months in educational age. When the test scores on the earlier forms of the test were translated into the new values, it was found that both the health instructed group and the control group tended to exceed the standards. At the beginning of the investigation, the average pupil among the cases under observation was in the advanced section of

Grade III and had an educational age of 9 years 11 months, which is two to four months in advance of the norm. In two years when the pupils were completing Grade V their educational age was 12 years 7 months, as contrasted with the expected standard of 11 years 8 months.

How much of this gain in test scores was the result of the previous experience with the test situation and how much was due to school improvement during the health program cannot be definitely determined from these data. However, it can be stated with certainty that the test gains of the health instructed pupils exceeded the usual standard for pupils from similar age and general intelligence levels, since the chances were less than 1 in 10,000 that their gain was as small as 25 months. The test gains of the two groups in educational achievement were as follows, using the New Stanford Achievement norms:

TABLE III

	BOYS		GIRLS		BOYS & GIRLS	
	Mean Gain	S.E.	Mean Gain	S.E.	Mean Gain	S.E.*
Health Instructed	33.5	± 1.85	32.4	± 1.62	33.0	± 1.24
Control	26.5	± 3.17	26.1	± 2.61	26.3	± 2.07
Expected Gain	25 mos.		25 mos.		25 mos.	

*The standard errors have not been corrected for correlation.

Comparisons of the achievement of the selected cases with the gains based on retest results from pupils in the same classes as the control group who took no part in the general program clearly showed that the health instructed pupils had made greater improvement than a non-instructed group who had had similar experience in amount of practice with the test materials. There were 171 pupils in the control schools who had no direct contact with any aspect of the health education program during the period of this research. These discards included children whose parents were uncertain as to their remaining in the community for the next few years; whose parents were uninterested in coöperating in a school health program; whose parents because of religious attitudes did not wish their children included in the health work; or whose parents understood too little English to participate in the plan. This non-participating group was not, then, an unselected one but their records provide a fairly accurate check on expected improvement under a yearly testing program.

It was not practical to translate all records from the pupil population into the values for the New Stanford Achievement norms. Hence, the

comparative findings in terms of the local standards are presented in terms of the 1926 norms. Table IV shows the amount of gain during 24 months in general educational achievement and in the special subjects of reading, arithmetic, and spelling. Without exception, the gains made by the pupils participating in the health projects were greater than those made by either the control group or the discards. On the whole, it seems highly probable that a true difference between the test scores of the health instructed group and of the discards at the same grade level was established, the critical ratio in standard error units being 3.00. In general, the rate of improvement of the control group appeared to be the same as that of their classmates who were not included in the investigation.

TABLE IV

IMPROVEMENT AS SHOWN BY SCHOOL TESTS OF THE SELECTED CASES CONTRASTED WITH THAT OF THEIR CLASSMATES WHO HAD NO CONTACTS WITH THE HEALTH PROGRAM (1926 NORMS)

	LOCAL NORM (Discards from Control Group 171)		CONTROL (104)		INSTRUCTED (268)		INSTRUCTED vs. NORM Diff.	CONTROL vs. NORM Diff.
	Mean Gain	S.E.*	Mean Gain	S.E.*	Mean Gain	S.E.*	S.D. diff.	S.D. diff.
Reading.....	22.1	1.47	24.0	2.24	29.4	1.48	3.52	.73
Arithmetic ..	23.7	1.39	25.5	2.54	29.1	1.25	2.87	.62
Spelling.....	22.2	1.22	17.2	2.02	23.0	3.73	.39	2.12
Educational Achievement	20.1	1.10	20.2	1.65	24.8	1.02	3.17	.05

*Uncorrected for correlation.

A criticism against the conclusion that the health instructed pupils actually improved more in their school work than pupils outside of the experiment is that the former were very probably more intelligent than the latter and therefore should have surpassed the latter's level of attainment. The health instructed pupils were 4.4 educational months more mature at the initial test and 9.2 months at the third test than the discards. Differences in learning capacity would have a cumulative effect on the school data but it is not likely that they could account for this marked difference in educational maturity at the final test between the instructed pupils and the discards. Results previously mentioned bear on this point. It has been shown that when the health-trained pupils and their control were of the same age and were equally bright and equally successful on the first educational test, that the health-trained pupils tended to show the larger gains in all school subjects. In general achievement it could be said that the chances of a true difference were

better than 99 to 1. The present analyses have indicated that the control group gained during this period at a rate similar to that of their classmates who had no contacts with the health program. It would, therefore, follow that the comparisons of the instructed pupils with the discards can be interpreted in the same way as the results from the instructed versus the control when variations in I.Q. were eliminated. From these findings it would seem unlikely that slightly higher intelligence ratings could account for the marked difference between the gains of the health instructed pupils and the discards. The inference is that participation in the health program was an influential factor in the improvement in school achievement.

Further evidence of the influence of the health projects upon educational achievement was revealed in the results relative to variations in the extensiveness of the pupils' contacts with the general program. Five gradations were contrasted. These ranged from no direct contacts of any type to a very broad health education schedule as follows:

1. Classmates of the control group who had no special contacts with any phase of the health work.
2. The control pupils who had a regular program of examinations and interviews.
3. Classmates of the experimental group who took part in the health instruction projects.
4. Children who transferred from the instructed to the control or vice versa, thus having a schedule of examinations, interviews, and a portion of the class instruction.
5. The group who had the complete program of examinations, interviews and health instruction.

The amount of gain in test scores during the period covered by this study is presented in Figure II in terms of the educational ages at the time of the school tests according to the extensiveness of the contacts with the health educational activities. These data are from the total groups irrespective of their initial standing. It is apparent from the graph that the gains made by the pupils participating in the classroom instruction were larger than those made by the non-participants. The chances were better than 90 in 100 that the pupils who took some part in the health projects conducted in the classrooms, whether they completed the schedule or had only two semesters of it, showed greater progress in their school work than those who had no contacts with the instruction. A true difference was established between pupils who had extensive contacts with the

health program and those who had very limited or no direct contacts with the program.

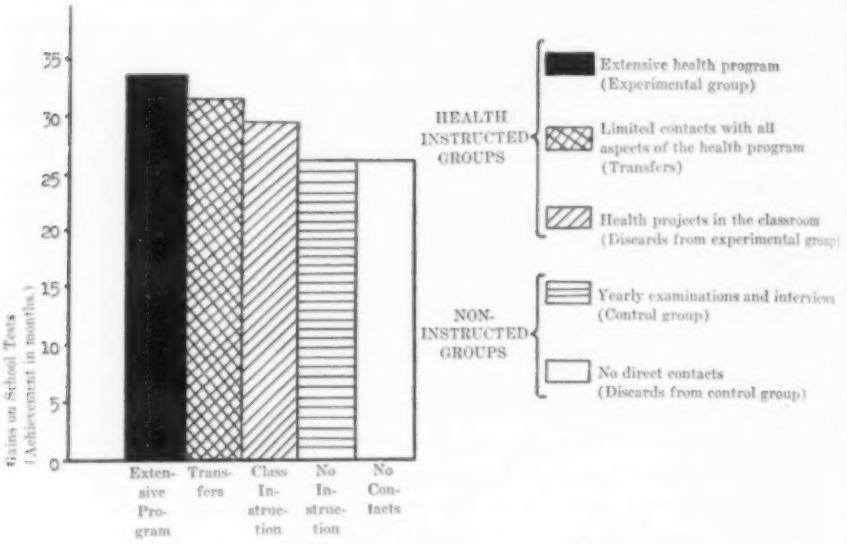

FIGURE II: SCHOOL IMPROVEMENT AS RELATED TO EXTENSIVENESS OF THE HEALTH PROGRAM

Teaching of health topics when combined with other types of health education appeared to be a greater stimulus than the instruction projects alone, although the differences were never large. The group who participated in a portion of the instruction schedule and who also had annual conferences with a physician, a health education specialist, and a psychologist, tended to make larger gains than a group who had only the classroom instruction. They, in turn, were exceeded by pupils who had the complete teaching schedule and regular interviews with the specialists.

The fact that group differences occurred does not establish a relation to the health education projects. It is possible that some special dissimilarities in the home background of the groups may have been largely responsible for the variation discovered. The factors which have been eliminated from the comparative results thus far have been those closely associated with the testing situation. Educational accomplishment is a complex problem involving the interaction of numerous conditions. Any attempt to isolate these conditions will, of course, tend to distort the total pattern but at the same time it will provide a check on the trend of the

present comparative findings. Certain environmental and social conditions were singled out for special study in this connection. These included nationality, socio-economic level and cultural background of the family, the type of activities which engaged the child's out-of-school hours, and his general attitude toward school work.

ACHIEVEMENT DURING THE HEALTH PROGRAM WHEN THE GROUPS WERE EQUATED WITH RESPECT TO GENERAL BACKGROUND

A careful examination of the comparative findings when several different measures of home condition were utilized in equating the groups revealed no exceptions to the general trend previously established. The increments of educational improvement were higher for the health instructed pupils than for the control pupils who were of like ages, intelligence, educational status, and home background. Table V briefly summarizes the results when the comparisons were based on pupils: (1) whose parents were native born and in whose homes English was used exclusively; (2) whose fathers were engaged in similar types of occupations as judged by Barr's scale of occupational intelligence; (3) whose mothers had had similar amounts of formal schooling; and, (4) whose homes were equally satisfactory with respect to material advantages. In every comparison the health instructed pupils, on the average, were found to have a somewhat higher rate of achievement, according to standard school tests, than the control pupils.

TABLE V

SCHOOL IMPROVEMENT AND HEALTH EDUCATION WHEN THE GROUPS WERE EQUATED

Factors Controlled	Monthly Increment of Educational Improvement			
	Health Instructed		Control	
	N	Mean	N	Mean
Nationality, age, I. Q., and E. Q.	122	1.09	40	.92
Father's occupational status, pupil age, I.Q., and E.Q.	160	1.03	63	.96
Mother's schooling, pupil age, I. Q., and E. Q.	147	1.02	58	.95
Home rating, age, I. Q., and E. Q.	135	1.04	52	.96

The activities which consume the child's out-of-school hours may serve to retard or to stimulate his general educational development. Particularly would this be noted if the child was fatigued and over-stimulated by too many interests or if through materials handled and the training received he gained special practice in the type of abilities and information measured by the tests. Inspection of the yearly records collected

from these pupils indicated that the groups were similar with respect to their play interests, their membership in organizations, the time given to leisure-time reading, and the frequency of attendance at the movies. None of the measures of interest in out-of-school activities appeared to be significantly related to pupil achievement. Among the population surveyed the highest coefficient of correlation for these activities in relation to educational accomplishment was found with respect to extent of leisure-time reading, in which r was only .29 ± .03.

One aspect of the reading data merits attention here. The most striking difference between the interests of the two groups was noted in the case of the amount of reading done by the girls in the health instructed classes. In the growth analyses (both mental and educational) the girls showed regular and consistent differences over their control while the boys in certain tests tended to be somewhat erratic. A connection between the reading and growth findings might be suggested were it not for the fact that the health instructed boys made larger gains on tests than the girls. Among the pupils participating in the health projects there were twice as many girls as boys who reported a large amount of leisure-time reading but on tests of reading facility and general school achievement the girls did not show as rapid progress as the boys, although they were equally bright. Thus there was no indication that reading interests operated to the advantage of the health instructed pupils.

A detailed inspection of the data on school attitudes likewise yielded no evidence of the influence of variations in pupil habits and attitudes on the observed group differences in school improvement. Differences in such personal attitudes as industriousness, coöperation, and persistence, as rated by the teachers during the period of the investigation, and in the general descriptions of their reaction time, as given by teachers and by the physician, appeared to be relatively independent of the test results.

SUMMARY AND GENERAL CONCLUSIONS

The 409 elementary school pupils included in this investigation were all physically able to be in regular attendance and could be said to form a representative sample of an average elementary school population. Their rate of educational development should have approximated the reported age and grade norms. On standardized tests participants in the health education projects exceeded the grade standards and one can be reasonably confident that these pupils were accomplishing more than is

commonly expected of the average pupil in his progress through the intermediate grades. The non-participants while up to standard gave no indication of an accelerated rate of achievement.

The consistency of the findings when the groups were alike at the beginning of the study with respect to age, general ability, school achievement, practice in the testing situation, and with respect to many aspects of home background and the general character of their school reactions, would seem to warrant the conclusion that the health education program, on the whole, had a stimulating effect on school achievement. However, there is no evidence that it was the health content of the teaching projects which resulted in the improvement. The entire set-up of the investigation, especially the phases involving conferences with teachers, principals, and parents, very likely served as a general incentive influencing teacher-attitudes, methods of teaching, school atmosphere, and, indirectly, individual accomplishment. It is believed to be a composite of all these extrinsic situations interacting with an improved health regimen and physical conditions—in both of which the instructed pupils showed marked improvement—that was associated with the larger gains in educational achievement of the pupils who received special health instruction.

THE COMPUTATIONAL ARITHMETIC OF SOCIAL EXPERIENCES OF THIRD GRADE CHILDREN

EBBA L. WAHLSTROM
Ohio University

"WHAT arithmetic shall we teach?" This was the question that Wilson[1] endeavored to answer in 1919. To do this he employed the school-pupil-survey method, i. e., the children were asked to report to the teachers problems confronting their parents in actual situations of everyday life. The findings indicated that the computations of the average adult were relatively simple. This fact was corroborated by similar studies made in Marshalltown, Iowa, in Boston, and in Melrose, Massachusetts.

Little has been done in the field of children's problems. By the means of personal interviews, Noon[2] endeavored to discover what arithmetical situations were encountered by boys in the intermediate grades of the Harry L. Pierce School of Boston. The technique of Smith[3] in her study of first grade children's problems was of the same type. Five hundred first graders in the schools of Detroit were questioned concerning their activities into which numbers or arithmetic entered.

The purpose of this study was to determine, in so far as it was possible to do so, the computational arithmetic involved in the every day experiences of third grade children. The investigation was conducted during the years 1930-1932 at the Arithmetic Research Laboratory of Iowa State University. The procedure followed the same general plan as was used by Wilson in his pioneer study. The services of thirty-one third grade teachers were secured. A fairly representative sampling of problems was obtained as the schools were located in the states of Louisiana, Oklahoma, Illinois, Iowa, Nebraska, Massachusetts, Pennsylvania, and New York. Three kinds of schools participated: public, parochial, and teachers college training schools.

The children were requested to report orally to the teachers their own actual experiences which involved arithmetical computations. These were stated in the form of problems and were recorded verbatim. In

[1] Wilson, G. M. *A Survey of Social and Business Usages of Arithmetic*, Contributions to Education No. 100, Teachers College, Columbia University, 1919.

[2] Noon, Philo G. "The Child's Use of Numbers," *Journal of Educational Psychology*, X (November, 1919), pp. 462-67.

[3] Smith, Nila B. "The Uses of Arithmetic in the Out of School Life of First Grade Children," *Elementary School Journal*, XXIV (April, 1924), pp. 621-23.

order to eliminate adult influences, teachers were cautioned not to accept problems brought from home in written form. By June 1, 1931, 1,687 problems had been contributed. The investigator felt that these were the bona-fide problems met with in the experiences of children, at least insofar as it is possible to secure such problems.

Detailed analyses of the problems revealed interesting facts. Which of the four fundamental processes was used most frequently by these children? Wilson discovered that adults employed multiplication more than they did the other processes. Table I shows what this study found in answer to the above. One notes that addition and subtraction are used almost to the exclusion of the other two processes.

TABLE I

FUNDAMENTAL PROCESSES USED IN THESE PROBLEMS

Processes	Frequency Distribution	Percentage Distribution
Addition	1005	48.7
Subtraction	796	38.6
Multiplication	223	10.8
Division	40	1.9

Further consideration of these same data showing the specific abilities found in each process, is presented in Table II. Apparently in the

TABLE II

SPECIFIC ABILITIES WITHIN EACH OF THE FUNDAMENTAL PROCESSES

Process and Specific Ability	Frequency Distribution	Percentage Distribution
Addition		
Basic facts (in isolation)	235	23.3
Adding by ending	20	11.9
Single columns	89	8.9
Two or more columns (no carrying)	336	33.4
Two or more columns (carrying)	213	11.3
Broken column	112	11.2
	1005	100.0
Subtraction		
Basic facts (in isolation)	152	19.0
Harder subtraction (no borrowing)	349	43.8
Harder subtraction (borrowing)	295	37.2
	796	100.0
Multiplication		
Basic facts (in isolation)	164	73.5
Harder multiplication	59	26.5
	223	100.0
Division		
Basic facts (in isolation)	23	57.5
Harder division	17	42.5
	40	100.0

actual experiences of these children multiplication and division are encountered mainly as basic facts, in isolation, such as 4×3. Addition and subtraction function in the harder or more complicated types.

How difficult are the problems that these children are called upon to solve? Table III shows that these children's actual arithmetical experiences are very simple. Ninety percent of all addition needed in these problems had but one or two-digit addends, while seventy percent had only two addends. Two-thirds of all subtracting done involved only one or two-digit minuends.

TABLE III

DIFFICULTY OF FUNDAMENTAL PROCESSES

Difficulty	Frequency	Percentage
Addition		
One-digit addend	322	32.03
Two-digit addend	620	61.19
Three or more	63	6.28
Two addends	703	69.9
Three or more	302	30.1
Subtraction		
One-digit minuend	85	10.71
Two-digit minuend	524	65.99
Three or more	187	24.0
Multiplication		
One-digit multiplier	216	96.8
Two-digit multiplier	7	3.2
Division		
One-digit divisor	22	55.0
Two-digit divisor (12 included)	18	45.0

A tabulation of the basic facts was also made. The social utility of the third grader's arithmetic, as shown by this study, spreads itself over all the basic facts in addition and subtraction while in multiplication the facts seem limited to the tables through the fives.

One step problems, as might be expected, were most common having a percentage frequency of 78.4. Fractions were used sparingly, while denominate numbers featured mainly as names of quantities and measures.

TABLE IV

AMOUNTS OF UNITED STATES MONEY USED

Amount	Frequency Distribution	Percentage Distribution
.01 to .25	280	32.7
.26 to .50	253	29.6
.51 to 1.00	182	21.3
1.01 to 4.00	95	11.2
4.01 and above	45	5.2

Of all problems contributed, sixty-one percent dealt with United States money. The amounts involved were comparatively small as Table IV reveals.

The final consideration was that of the activities which featured in these everyday experiences of third grade children. Buying had the highest frequency. A detailed analysis showed that children purchased all manner and sort of things. It was interesting to note that they bought more for other people than for themselves. Work situations included selling papers, errands, caring for animals, housework, etc. Games showed an equally great variety.

TABLE V

ACTIVITIES AND TOPICS

Activity	Percentage Distribution
Buying	34.66
Counting	17.07
Work Situations	11.49
Games	8.53
School Activities	9.66
Amusements, (Shows, Circus, etc.)	4.97
Gifts	3.49
Savings and Allowances	2.45
Winning and Losing Things	1.89
Understanding of Adult Activities	1.41
Construction	1.24
Health	.94

The data obtained from this study reveal departures from current practices. Some students of the curriculum question the making of a course of study based on the natural social environment of children. It is of interest, however, to compare the information secured from this survey with an Arithmetic course of study suggested by one group of curriculum makers.[4] The course, as found in Fourth Yearbook of Superintendence is given below:

III Grade

The emphasis in this grade is on addition and subtraction.

1. Addition (primary and decade through 80) and subtraction (up to 9 from 18) mastered. Teaching children to check or prove their own work.

[4] *Fourth Yearbook*, Department of Superintendence, National Education Association, 1926. (Chapter on Arithmetic.)

2. Multiplication to 5×5 (31 facts) as time permits.
3. Counting by 1, 2, 5, and to 100.
4. Reading and writing numbers up to 1000.
5. United States money used with decimal point.
6. Fractions, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{4}$, $\frac{3}{4}$, used and understood—simple addition and subtraction objectively as time permits and needs arise.
7. Measuring and weighing in units as mentioned in second grade with the pound, quart, gallon, or new units.
8. Life situations in buying.

Similarities are to be noted. The emphasis placed upon addition and subtraction in third grade by these students of the curriculum coincides with the emphasis as placed by the result of this survey. The recommendation as to the amount of multiplication also tallies with that of the children's use of that process in this set of problems. Practically no division occurred in this study. Neither is division, as division, included in this outline for third grade.

A comparison of the findings of this investigation with present day courses of study shows a lack of agreement as to which of the fundamental processes receive the predominate stress. If new and recently revised editions of Arithmetic texts may be regarded as an index of what is taught in the third grade, one often finds that a large or major portion of the time is devoted to the processes of multiplication and division. The small use made of these processes in this collection of problems presents the question as to the reasons for this existing situation. Three possibilities might be suggested: (a) children may have avoided these types of problems because of their difficulty; however, they were not asked to solve their examples; (b) children may not have recognized these situations as problems and if so, they possibly have no real need for them at this age level; (c) children may not have reached a stage of development mature enough for solving these types of problems, as the processes of multiplication and division may be much more difficult than they have been considered in the past.

The question of allocation of subject matter on the basis of the child's needs and maturation rather than on the basis of logical determination sanctioned by tradition is one which should be open to a great deal of experimentation and study. May not many of the arithmetical functions, which are deemed simple and easy by the adult person, be the reverse for the child who has no immediate or constant need for them? How valid a basis of curriculum making in arithmetic are the current

needs of number that arise in the types of activities in which children now ordinarily engage?

The following are the conclusions to be drawn from this survey:

1. These practices exist in regard to the use of the four fundamental processes in the activities of third grade children:
   (a) Addition and subtraction are used almost to the exclusion of multiplication and division.
   (b) All types of addition and subtraction were encountered, while in multiplication the basic facts in isolation, as 4×3, had the highest frequency of usage.
   (c) The analysis of the difficulty of the fundamental processes as shown by size of the addends, the minuends, the multipliers, and the divisors indicated the relative simplicity of these computations.
   (d) The arithmetic of social utility spreads itself over all the basic facts in addition and subtraction, while in multiplication the facts are limited mainly to the tables through the fives.
2. Seventy-eight and four tenth's percent of the problems required but one step in the solution of them.
3. Money transactions are highly important in the child's every day experiences even though the denominations used are small.
4. The activitiy analysis indicated buying as the one having the greatest frequency of occurrence.
5. The arithmetic of social experiences of third grade children as shown by this survey would possibly suggest further simplification of current practices found existing in the present arithmetic curriculum for the third grade. On the other hand the question may be raised, should not the school undertake to extend and to enrich the number experiences of children?

EDITORIAL

TRAINING FIELD WORKERS AND GRADUATES IN METHODS OF RESEARCH

With respect to providing training in research methodology, there are prevalent in education at least two points of view,—points of view that are not necessarily mutually exclusive and often are complementary. In the conventional longer established approach to training in research the instruction is supplied through special seminars, the laboratory, and research courses. At times these are supplemented by courses in statistics, measurements, history and survey techniques. By this plan such familiarity with investigational methods as was acquired came more or less incidentally through contact with advanced courses in specific educational areas. All of these avenues offer, of course, genuine possibilities for a functional understanding of research procedure, though the emphasis may too often have been placed on the results or data of the problem investigated to the partial neglect of a real grasp of scientific method as a method of problem solving. A second more recently introduced method of training research workers is the introductory survey course in the Methodology of Educational Research, which may accompany or precede the specialized seminar.

When it is suggested to professors or administrative officers holding the earlier point of view that first year graduate students need fairly broad introductory or orientation training in the methodology of research before undertaking one or more specialized courses in research technique or one or more seminars in a limited area of education, the answer sometimes has been that: (1) not enough material is available for a general course, (2) such a course is not sufficiently specialized to facilitate an attack on definite problems, (3) students can do such reading independently for themselves, (4) the professor can give the necessary information through conferences with the thesis candidate, or (5) research methods can best be illustrated and described in connection with other courses,—for example, administration, psychology, and history of education classes may emphasize respectively the survey, experimental,

and historical techniques of investigation. Discussion of these points is reserved for later comment in this editorial.

In the general survey course the steps in problem solving and the functional relationships between the various investigational approaches appear in a pattern of research methodology, not likely to result from participation in only one or two specialized research courses or seminars. Appropriate topics for consideration are: principles and illustrations dealing with characteristics of scientific method, selection and delimitation of the problem, bibliographical and library techniques, the various research procedures or approaches for collecting data, analysis and interpretation of data, and the preparation of the research report. It is entirely appropriate that seminars also cover such a sequence of topics in conjunction with the individual problems or theses, provided similar training has not been secured in a general introductory course, with provision for placing a comprehensive body of reading materials in the hands of the student as a ready guide and reference work. Whatever the type of course or seminar in which the worker receives his training in the methodological steps and relationships of investigational procedures, the functional or activity approach in connection with a problem or a thesis well may be emphasized rather than to employ a formal textbook plan with emphasis only on the learning of a body of subject matter.

To suggest some of the merits of this second plan: (1) the accumulated printed material on research methods of the past ten years is ample for a good general course, which may well include an introduction to the literature of research; (2) it does not pretend to take the place of other specialized courses on research techniques, but enables the student to make an intelligent choice between such courses in relation to his problem and needs, as well as provides familiarity with a wide range of methods, some of which are outside the knowledge of the student who has taken only one or two specialized research courses; (3) the reading and study of students on general methods of investigation are organized in a definite, systematic manner rather than left largely to chance; (4) professors are relieved of the unnecessary burden of imparting repeatedly in individual conferences the information presented in the general course; and (5) the applications pointed out in the courses dealing with specific areas of education should be more meaningful by reason of the students' general background in investigational procedures.

As a matter of fact, the fundamental principles, relationships, and steps in research, as represented by the topics suggested above for a gen-

eral course, should be understood by all competent educational workers in the field or on the campus (whether graduate students, administrators, supervisors, or teachers), if they are to contribute most effectively to the solution of problems in their respective areas of responsibility: (1) all as consumers of research seeking to make appropriate applications; (2) many as active participants in the problem-solving approach because of the resulting training value and the increased understanding of the educative process; and (3) a limited number of actual producers of fundamental research who contribute in some significant way to the existing store of educational knowledge. The extent to which individual workers go beyond the introductory foundational training in methods of research into more highly specialized technical courses in research (concerned with a single approach such as a testing, statistical, survey, historical, experimental, interview, or case procedure) must be determined by the needs of the given institution, group, individual, or school system.

CARTER V. GOOD

REVIEWS

Address all communications relative to book reviews to Editor Kai Jensen, University of Wisconsin, Madison, Wisconsin.

PETERS, CHARLES C., AND VAN VOORHIS, WALTER R. *Statistical Procedures and their Mathematical Bases* (State College: The School of Education, Pennsylvania State College, 1935), pp. vii + 363. $1.50.

"This book is printed by the photo-offset process, by which process photographic negatives are made directly from the typewritten page, and positives from these on thin metal plates;" thus the expense of typesetting is avoided. In order to avoid the uneven right-hand margin, characteristic of typewritten material, Mr. Peters, invented a mechanism which makes all lines the same length. In using the mechanism, the typist notes the number of spaces in excess or deficiency in a given line and adjusts the mechanism accordingly before beginning the line.

The book is an effort "to take the magic out of statistics" by giving mathematical developments "in such form that persons who have studied only high school mathematics, plus the chapter on calculus with which this book opens, can follow them."

There is unquestionably a definite need for a book of this type which presents derivations in greater detail than is customary in more advanced texts so that beginning students will have a convenient source for learning the mathematical theory and assumptions behind the formulae. Unfortunately, however, this book, while presenting certain derivations very accurately and clearly, presents others which are definitely erroneous. A good part of the material is standard elementary statistics, such as the presentation of mean, mode, median, standard deviation, other measures of variability, the Pearson r, the correlation ratio, and methods of computing these statistics. This is standard material, clearly presented and as such calls for no special comment. Since the principal emphasis of the work is on mathematical bases, the derivations presented should be considered in more detail. A number of these derivations are presented very clearly and accurately, for example, the explanation of the constant e, bi-serial r, and the bi-serial and tetrachoric r when the central portion of the distribution is unknown, the correction of coefficients of correlation for broad categories, and curve fitting by least squares as applied to the straight line, parabola, exponential, and Gompertz curves. Correlation theory is suitably introduced by a consideration of the regression line. The derivations in the chapter on correlation are presented very clearly and accurately. The chapter on reliability of statistics presents an interesting series of standard errors: first, the simple standard error of a statistic; second, the standard error of a statistic in the case of correlated samples matched on a fallible criterion; and third, the standard error of a statistic in the case of correlated samples

matched on an infallible criterion. These formulae are given for the standard error of a mean, of a standard deviation, and of the difference of two means. The derivations in chapter VII, Inferring Coefficients of Correlation for Changed Conditions, are well presented in general. The chapter begins with a derivation of the formula for a correlation of sums or averages. From this the Spearman-Brown formula, the correction for attenuation, and the correction for heterogeneity follow readily. A good presentation is given of partial and multiple correlation although the more condensed detrimental solutions are not given. The chapter on the normal curve contains a very good derivation of various properties of the normal curve, such as the mean deviation of a portion of a normal distribution, the values of B_1, B_2, and the inflection points.

If the standard of clearness and accuracy shown in the sections mentioned above were maintained throughout, the book would constitute a valuable addition to statistics textbooks. Unfortunately, however, the book, as originally set up, contained a large number of typographical errors, about one hundred and fifty of which have been corrected in supplementary four page folder. Copies of the book now sold have had these corrections inserted by hand, but in the review copy some of these corrections are also faulty. In addition, there are still some fifteen fairly serious typographical errors in formulae or in mathematical symbols in the running text, errors serious enough to confuse "persons who have studied only high school mathematics," such as: x, instead of Δx, (page 15); $\int (x)\,dx$, instead of $\int f(x)\,dx$, (page 27); f (x) instead of f (a), (page 68); ΣD and ΣD^2, instead of Σn and Σn^2, (page 89); and $N\sigma^2$, instead of $N\sigma^4$, (page 120). An exponent, n, is omitted in formulae 20 and 21 on page 246; several terms are omitted in the formulae on pages 271 and 280. On page 66, in discussing a positively skewed distribution, the statement is made "its mean will be larger than its median and its mode will tend to lie between these two." The statement would be correct if the words "median" and "mode" were interchanged. On page 317 we find "L. L. Thurstone has found that the parabola seems best to fit the norms of forty tests and claims this type to be the curve of learning." The curve used by Thurstone was the hyperbola.

All formulae are typed, which is quite confusing particularly in the case of formulae which contain fractional exponents and coefficients, such as the equation for the normal curve and its first and second derivatives. (See pages 23 and 236.)

In addition to the errors listed above which, although confusing to a beginner, can be regarded as typographical errors, there are also a number of rather serious errors in statistical and mathematical logic. The authors, for example, confuse "kurtosis" with "variance," suggesting in discussing the coefficient of variation (on page 65) that "a much more promising standard measure of variability comparable for all distributions would be the measure of kurtosis." Taking this statement literally would lead one to conclude that all normal distributions have the same variability, which is obviously untenable. The derivation of the standard error of a difference between means contains two compensating errors by means of which a correct result is obtained. The two formulae on page 140 should contain an additional factor $\frac{Sn}{N}$. This error is compensated for by the erroneous substitu-

tion $Sn = N$ on page 141. Actually, as the derivation is set up, $S = \frac{N!}{n!\,(N-n)!}$. On page 162 it is assumed that the geometric mean of two correlation coefficients may be substituted for the arithmetic mean, but this assumption is not stated in the derivation. In chapter XII, The Technique of Controlled Experimentation, we find "If the tests employed have low reliabilities, on account of shortness or other limitation, the obtained differences are smaller than the true ones," (page 330). This conclusion is erroneous. The best assumption to make regarding mean error is that it is zero, which makes the mean obtained difference equal to the mean true difference. The obscuring effect occurs only when the standard error of the differences is also considered. A correction can be found by using the true variance of the test instead of the observed variance in the formula for the standard error of a difference. Considering the obscuring effect of the lack of validity in a test, the authors derive a formula which states that the true difference is equal to the observed difference divided by the validity coefficient of the test. This is not generally true but holds only under certain special assumptions, which are not stated in the derivation. On page 339 the authors derive a formula,

$$r_{23} = \tfrac{1}{2}\,(r^2_{12} + r^2_{13}),$$

for estimating the correlation between two variables from the correlation of each with a third variable, when the two are entirely uncorrelated except through their correlation with the third variable. Their assumption that the variance of a single column of differences is equal to the variance of all the differences is true only if there is zero correlation between the differences and the matching variable. The correct formula, $r_{23} = r_{12}\,r_{13}$, can readily be derived from the partial correlation formula.

The examples of uncorrected errors given above would seem to show that the book as it stands cannot furnish a safe and convenient source of *derivations* for persons with only an elementary knowledge of statistics and mathematics. If, in a future edition, the authors correct the errors, the book will then fill a definite need in the field of elementary statistics. The final formulae given throughout the book are correct (except for those cited in chapter XII) so that as a source of statistical formulae the book is reliable.

HAROLD GULLIKSEN

University of Chicago

WILLS, ELBERT VAUGHAN. *The Growth of American Higher Education: Liberal, Professional and Technical* (Philadelphia: Dorrance and Company, 1936), 225 pp. $2.00.

Wills' *Growth of American Higher Education* is a brief treatment of the various types of higher institutions: colonial, state, denominational, professional and technical, teacher-training, women's, endowed, graduate, urban, and junior college. The brevity of the discussions may be illustrated by reference to the 13-page chapter on the higher education of women, whereas Thomas Woody has devoted two volumes to this subject. The documentation is light, only some 45 citations being made. The bibliography includes only 44 references, with some significant omissions such as Koos' work on the junior college and Reeves and Russell's surveys of denominational and liberal arts colleges. The book presents a convenient brief summary or overview of the development of higher education, although the treatment is too sketchy to render it satisfactory as a textbook or reference

volume for students of the history of education.

CARTER V. GOOD
University of Cincinnati

WALLIN, J. E. WALLACE. *Personality Maladjustments and Mental Hygiene* (New York: McGraw-Hill Book Company, 1935), pp. xii + 511. $3.00.

From first impression to the end of a careful reading the numerous personal incidents and case presentations caught and held the reviewer's attention. He is inclined to the opinion that time will prove these to be the most useful part, which is a very large part, of this book. The author thinks of mental hygiene as a field in which material from many other sciences may be applied. He says that for mental hygienic purposes it is necessary to draw upon psychology, child study, education, sociology, psychiatry, medicine, and biology. Two problems are in the author's conception involved. One of these is the "preservation of the mental health of the individual and of the community." The other is the prevention and cure of mental disease. This inclusion of cure of mental disease as a part of mental hygiene is noteworthy. Many would not agree to this. So conceived a text on mental hygiene must of necessity cover a very wide range of subjects and this one is no exception. The subject matter includes such diverse fields and problems as those of mind-body relationship, therapeutics, characteristics of the well-adjusted person, content of a mental-hygiene program, conditioning, inferiority troubles, value of humor, rationalization, regression, day-dreaming, hallucination, drives, conflicts, psychoanalytic theory, dissociation, control of fear.

The style and organization impressed the reviewer as unusually trying. Ordinarily an author has a general pattern or plan worked out before approaching the task of actual composition and throughout his presentation this plan dominates. But in this instance the discovery of the author's plan constituted for the reviewer a continuously baffling task. The style itself varies from what appears to be a stenographic report of informal presentation to that of careful composition. And there is much repetition. In the preface the author tells his reader that the material of the book "first appeared in the form of lectures . . . offered both to popular audiences and to students." It is quite evident that he was also conscious of the repetitiousness of his work for he defends it on the ground that his students have found the much repetition "distinctly advantageous." Perhaps those students were right but the reviewer would have enjoyed less of the repetition and a more systematic organization. He also doubts the wisdom of introducing such technical terms as conflict, phobia and hysteria prior to their definition.

It is evident that behind the writing of this volume there was both a large personal experience with human behavior problems and a wide knowledge of the literature. The reader is given richly of the fruits of that personal experience with behavior problems; but not so profitably of that knowledge of the literature. There is an extensive bibliography at the end, but many allusions are made in the course of the text to the work of various persons whose names are mentioned but quite without authenticating references. The bibliography is notably defective from the point of view of the careful bibliographer. In the listing of books the place of publication and the name of the publisher and the number of pages are omitted. Journal ref-

erences are given without the volume number. These bibliographic imperfections, however, are more likely the fault of the publisher. When will publishers learn the usefulness of such items!

A few curious and unfortunate omissions may be noted. Speech defects are included among the causes of personality conflicts and behavior disturbances. But there is no insightful presentation of the psychology of speech defects. Surely a mental hygiene must grow out of an understanding of the nature of the causes of mental disturbance. Blindness also appears as a cause but of the literature on blindness now so copious nothing is given. The only-child situation is listed as a common factor in the production of mental difficulties. No references in support of this contention appear and the reader is left without an introduction to a literature which is now very useful. In the presentation of the psychoanalytic conception of the stages of love development both the narcissistic and the parental stages are omitted. Why this should be is especially puzzling. Narcissism is today presented by many psychiatrists as the essential feature of schizophrenia. Elsewhere in the text the author himself makes much of the disturbing effects of parental fixation, hence it must be that he accepts the parental stage as one in the course of normal development.

EDMUND S. CONKLIN
Indiana University

MACGREGOR, GREGOR. *Achievement Tests in the Primary School. A Comparative Study with American Tests in Fife.* Publication of the Scottish Council for Research in Education, No. 6 (London: University of London Press Ltd., 1934), pp. xiii + 136. 5 shillings.

One often hears from the public platform or encounters in professional or popular reading comparisons of the relative efficiency of the American and European school systems, but rarely do those making the comparisons have objective facts upon which to base their judgments. Consequently, educators and the professionally-minded public should welcome this volume which presents the results obtained from administering Battery A of the Public School Achievement Tests in the public schools of Fife to 6,265 pupils whose eleventh birthday fell in the session of 1930-31. This battery, including tests in reading, arithmetic computation, arithmetic reasoning, language usage and spelling, was standardized for use in America, and, as pointed out by the author of the volume, was much better suited for use in the American than in the Scottish schools. But in spite of the limitations, the tests were given and scored according to the instructions designed for use in America. In general, the results show that the level of achievement for the Fife "eleven-year-olds" is sixteen months superior to that for children of the same age in the public schools of America. Children in Fife are superior on the five tests as follows: reading, five months; arithmetic computation, twenty months; arithmetic reasoning, thirteen months; language usage, twenty-four months; spelling, twenty-nine months.

In an effort to account for the superiority of the pupils in the Scottish schools the author, after careful and elaborate statistical analysis of the results obtained from the administration of the tests to the eleven-year-olds and from supplementary data secured through administering other forms of the same tests and the National Intelligence Tests to 1,326 pupils fairly equally distributed among ages 7 to 11, inclusive, in nine representative schools of Fife, concludes that there is

substantial evidence to show that the educational advantages which the pupils of Fife have over the pupils of America at eleven years of age is mainly due to their earlier enrolment in school. The author points out that the pupils of Fife, entering school at about five years of age, attain a level of achievement at the age of seven and one-half years which is fourteen months above that of the pupils in American schools and that they maintain this superiority in all subsequent ages. He considers such factors as intelligence, longer school day and year of the Scottish schools, the quality of the teaching staffs and the entering ages of the pupils, and decides that the vital factor is the fifteen months earlier entrance age.

This volume is worth reading. It is well written and presents sound techniques. It contains an excellent discussion of the limitations of the tests and of the difficulties involved in the efforts to compare standards of achievement based upon tests designed for use in a restricted area. The statement of conclusions, however, constitutes the challenging contribution. While American educators who are inclined to postpone the teaching of five subjects and processes until greater mental maturity of the pupils is reached should give serious consideration to the findings, the writer is not alarmed by the situation presented. He tends to believe that superiority may be explained through the wider range and polyglot nature of the American pupil population, the selection of schools in which the tests were standardized, and the differences in the emphasis on educational values in the schools of the two countries. He believes that the present comparisons are interesting but that much more intensive investigations must be made before definite conclusions can be established.

CLIFFORD WOODY
University of Michigan

UNZICKER, CECELIA E. *An Experimental Study of the Effect of the Use of the Typewriter on Beginning Reading*. Contributions to Education, No. 610 (New York: Bureau of Publications, Teachers College, Columbia University, 1934), pp. viii + 95. $1.50.

During recent years the use of the typewriter in school work has spread rapidly. The tendency to introduce typing into the lower levels of the elementary school has raised the question of its effect upon progress in other school subjects, especially in reading. In this monograph the author investigated the "effects of the use of the typewriter on learning to read by first grade pupils . . ."

In a carefully planned program an experimental group spent 75 to 90 minutes per week at typing while the control group was engaged in regular activities of the school. The experiment lasted 7 months with two measurements of reading progress. The results showed a slight but consistent trend of superiority for the typewriter group in all measures of reading achievement. The differences were not statistically significant however. Analysis revealed that only children with I.Q.'s of less than 90 profited uniformly from the typewriter training. There was no consistent trend at the higher levels of intelligence. The author fails to employ these findings to suggest that more emphasis upon the consecutive left to right sequence of individual letters in words might help the dull pupils greatly in learning to read. One may seriously question whether the contemporary methods of teaching reading are uniformly suited to all levels of mental ability.

This monograph has made a worthwhile contribution in showing that typing does not hinder learning to read in the first grade. The results should be checked with larger groups, however, before advocating the introduction of typewriters

into the primary grades as a general educational policy.

MILES A. TINKER
University of Minnesota

WADE, J. THOMAS. *A Measurement of the Secondary School as a Part of the Pupil's Environment.* Contributions to Education No. 647 (New York: Bureau of Publications, Teachers College, Columbia University, 1935), pp. vi + 68. $1.50.

Statistical technique applied to educational organization. Scales have been developed in recent years for rating educational organization, of which the Mort-Hilleboe scale for rating elementary school organizations is probably the most familiar. The techniques used have, in the main, been similar to the one used earlier by Strayer and Engelhardt in evolving their scales for the measurement of school buildings. A recent attempt to build a rating scale introduces an interesting variation and refinement of earlier methods. The author takes a list of objective questions intended to cover many of the factors in a secondary school. The questions, after going through various stages of refinement, number two hundred twenty-nine. These were submitted to principals of selected schools for rating. A preferred answer was selected for each question and each school was scored on the basis of these preferred answers. A random sample of papers was then taken and the biserial coefficient of correlation between each question and the total score was computed for all papers in the sample. Similar computations were made for three samples for the purpose of assuring a more suitable scoring key. The resulting key was used to score a sample of five hundred papers. From these scores the correlation of each of the questions with the total score of the papers was obtained. The questions were then classified into categories and those with biserial correlations of .30 or less were dropped. A list of one hundred questions survived this elimination process. This was used as a long scale for measuring secondary schools. A short scale which appeared to be about as reliable as the long one was subsequently developed. The usual methods of testing reliability were applied. The correlation between halves of the short scale was .89 which became .94 when corrected by the Spearman-Brown formula. Corresponding correlations for the long scale were .90 and .95. The scales were then tested by distributing the scores of five hundred papers on the long scale and of six hundred twenty-eight papers on the short scale with the usual result that both curves approached the normal distribution and that large schools scored consistently higher than small ones.

As a result of the study the author has derived a scale which can lay claim to being the most statistically refined of any yet derived. The basis of selection of the questions was the biserial correlations as derived by the aid of Table 3 of Thorndike's I.E.R. Statistical Tables, Series of 1925 in which "each entry is the coefficient of correlation corresponding to the percent of overlap stated in the same line, when the total group displays failures and successes in trait 2 in the proportions stated at the top of the page above the entry." To quote Thorndike:

"This table may be used in place of the calculation of biserial r where the percent of failures is between 10 and 90, with only slight losses in precision and reliability. . . . We have not determined the unreliability of an r computed by this table; but a reasonable procedure will be to use 1.3 times the error for biserial r as computed by the regular method."

That the scale measures whatever it measures reliably there can be little doubt. Two questions remain: Whether or not the assumption underlying the making of the scales, either in whole or in part, is entirely defensible will be left to the judgment of the reader. This assumption, as stated in the first paragraph of the study, is as follows:

"As the secondary school is a factor in the environment of each of the pupils attending it, each of its elements will exercise some influence upon every pupil. In the study, a number of these elements have been considered, and from them have been selected the ones thought to be related to the entire situation." The second question which each reader must also answer for himself as he carefully reads the one hundred questions, relates to the validity of each of the questions and to the validity of all of the questions taken as a complete measure. That is, does the total score of the scale represent a valid measure of the total environment? Are the questions which the author "thought" were related to the entire situation sufficiently related to justify the use of such scales as dependable measures of a *total* school of environment? There need be no doubt about their measuring *parts* of this environment. For the study to be complete, the judgments of some ten or twelve secondary school educational theorists with respect to the two questions raised herewith might be worth obtaining. The author has produced a meritorious scale which can be used to compare high schools in certain regards whenever such comparisons appear to be needed.

J. M. HUGHES
Northwestern University

RESEARCH ABSTRACTS and BIBLIOGRAPHIES

Address all communications relative to research abstracts and bibliographies to A. S. Barr, University of Wisconsin, Madison, Wisconsin.

BULLETINS

BROOKS, FOWLER D., ET. AL. *Abstracts of Papers At the St. Louis Meeting, 1936*, Yearbook XXIV of The National Society of College Teachers of Education (Chicago: The University of Chicago Press), 64 pp.

Contains summaries of papers relating to the improvement of college teaching and the plans for the 1937 Yearbook on the use of background in the interpretation of educational issues.

CAMPBELL, WILLIAM GILES. *A Form Book for Thesis Writing* (Revised edition; Los Angeles: University of Southern California Press, 1936), 145 pp.

Following the introduction the author discusses the format of the thesis; reference; tables, graphs, and illustrations; typing aids; matters of style; and Bibliography.

CARR, WILLIAM G., AND OTHERS. *The Preparation of Teachers Salary Schedules, Part II: Drafting the Schedule.* Research Bulletin of the National Education Association, XIV, No. 2 (Washington: Research Division of the National Education Association, March, 1936), 98 pp.

This bulletin discusses such matters as classifications of salaries, minimum salaries, maximum salaries, increments, professional requirements, sex, teaching efficiency, and other factors in salary schedules; and methods of putting a salary schedule into operation.

CHAMBERS, M. M., ET. AL. *The Fourth Yearbook of School Law*, (Washington: M. M. Chambers, 744 Jackson Place), 154 pp.

Contains materials relating to pupils and parents; teacher certification, contracts, etc., teacher tenure; local school officers; the acquisition, use and control of real property; school contracts; the tort liability of school districts; school district indebtedness; taxation; and related topics.

CRAMER, JOHN FRANCIS, *Australian Schools Through American Eyes.* Educational Research Series, No. 42 (New York: Oxford University Press, 1936), 59 pp.

The author gives his impression of the Australian schools, discussing the central organization and administration, the Australian teacher, inspection and supervision, secondary education and state examinations, miscellaneous observations and struggles between progress and tradition.

DOWELL, ANITA SHEMWELL. "The Physical Disability of Teachers in the White Elementary Schools of Baltimore, Maryland," *The John Hopkins University Studies in Education, No. 24,*

(Baltimore: The John Hopkins Press, 1936), 98 pp.

After reviewing previous investigations, stating the problem, and outlining the procedure for her investigation, the author presents data on the amount, frequency and duration of illness; the relation of certain factors, such as the month of the year, age, salary, etc. to the illness; the relation of promotion and marital status to illness; conclusions and recommendations for further studies. Bibliography.

DUNLAP, WILLIAM C. *Quaker Education in Baltimore and Virginia Yearly Meetings with an Account of Certain Meetings of Delaware and the Eastern Shore Affiliated with Philadelphia,* (Philadelphia: University of Pennsylvania, 1936), 574 pp.

An historical study tracing the origin and development of Quaker education in Maryland, Virginia, Delaware and adjacent parts of Pennsylvania, from the first settlement of Quakers in the latter part of the seventeenth century to the present day.

FARLEY, EUGENE SHEDDEN. *A Test of Factual Relations in American History* (Philadelphia: The University of Pennsylvania, 1935), 145 pp.

The author discusses the need for such a study, the determination of the content of the text, objectives of American History, instruction of the text, standardization of the text and the use of the text. Bibliography.

HECK, ARCH O., AND OTHERS. "Pupil Personnel, Guidance and Counseling," *Review of Educational Research, VI, No. 2* (Washington: American Educational Research Association, April, 1936), 275 pp.

Contains summaries of research relating to school attendance, school progress, school marks, recording and reporting, characteristics of pupil population, guidance and counseling, extra-curriculum activities, educational adjustment, special schools and classes and child labor. Bibliography.

KREY, AUGUST C. A City that Art Built. (Minneapolis: The University of Minnesota Press, 1936), 51 pp.

This is the second of a series of publications relating to the Owatonna Art Education Project, reviewing the history of the art of Florence and emphasizing how art came into existence out of fundamental life needs.

HERTZBERG, OSCAR E., AND MAYER, INA MEAD. *An Analysis of the Report of the New York State Teacher Training Institutions To Improve Their Product Through Better Selection* (Oscar E. Hertzberg, Head of Department of Psychology, State Teachers College, Buffalo, N. Y.), 60 pp.

The authors discuss the history of the selective admission program from 1932 through 1935 in New York state, the tests applied to freshman and advanced credit candidates and the results obtained.

SIMON, DONALD LIONEL. *Turnover Among Teachers in the Small Secondary Schools of Indiana,* (Chicago: The University of Chicago Press, 1936), 157 pp.

After stating the problem and the sources and methods of collecting information the author discusses the rate of turnover and length of tenure, major causes of teacher turnover, turnover for professional reasons, turnover for politi-

cal reasons, turnover for personal and other reasons, teacher qualifications and turnover, summary and implications. Bibliography.

SOHL, CHARLES EDWIN. *State Control of the Location, Planning and Erection of Public School Buildings* (Philadelphia: The University of Pennsylvania, 1936), 373 pp.

The author discusses the agency of control, school building sites, plans and specifications, bids, contracts, contractors, bonds, mechanics' liens, status of the architect and similar problems relating to the location plans and erection of public school buildings.

EDUCATIONAL ADMINISTRATION AND SUPERVISION

ANDERSON, EARL W. "Teaching Opportunities in 1935," *Educational Research Bulletin* (March 18, 1936), pp. 72-76.

The author presents an annual summary of the teaching opportunities in the state of Ohio.

ATKINS, BLANCHE E. AND ATKINS, RUTH E. "A Study of the Honesty of Prospective Teachers," *The Elementary School Journal*, XXXVI, (April, 1936), pp. 595-604.

The author presents the results of an experimental study of the honesty of teacher's college students from which they conclude that the proportion of students who are dishonest in tests is in direct relation to the procedure of the teacher.

BABB, A. S., AND JAYNE, C. D. "The Use of Sound Recording Equipment in the Study and Improvement of Teaching," *The Journal of Experimental Education* (March, 1936), pp. 279-287.

The authors present photographs and technical details relative to a sound recording instrument employed by them in the study of classroom instruction.

SMITH, E. L. "A Critical Analysis of Rating Sheets Now in Use for Rating Student-teachers," *Educational Administration and Supervision*, XXII (March, 1936), pp. 179-189.

The author presents an analysis of the rating sheets used by 103 colleges and universities, members of the North Central Association of Secondary Schools and Colleges.

WATKINS, RALPH K. "The Preparation of High School Science Teachers in Terms of the Market," *Science Education* (April, 1936), pp. 56-60.

The author offers data on the patterns of science combinations offered in First-Class Missouri High Schools for the school year, 1934-35, concluding that we still continue to train specialist for jobs that no longer exist.

WILKINS, WALTER L. "High School Achievement of Accelerated Pupils," *The School Review*, (April, 1936), pp. 268-274.

The author presents the results from a study of 282 accelerated pupils in a mid-west city, concluding that when acceleration is practised wisely, i. e., with careful individual study including appraisal of physical, mental and educational growth, emotional stability and social maturity, the results from the standpoint of intellectual adjustment of the accelerated pupil will, in general, be good.

PSYCHOLOGY OF TEACHING AND LEARNING

DOUGLASS, HARL R., AND PEDERSON, KENNETH L. "An Experimental Evaluation of A Unit Procedure in Teaching American History, *The School Review* (May, 1936), pp. 362-372.

From an experimental study of eight sections in American History in the high school at Hibbing, Minnesota, the authors conclude that for the unit plan of instruction employed in this experiment in the hands of well-trained teachers familiar with the plan the results secured are superior to those obtained by study-recitation plan of supervised study when measured by the use of objective tests.

PRATT, CARROLL C. "Repetition, Motivation, Recall," *The British Journal of Psychology*, XXVI (April, 1936), pp. 425-430.

The author presents the results of an experimental study of the presentation of a list of 100 two-digit numbers to 193 students with no repetition, and with motivation from which he concludes that maximal results are secured when repetition is supported by a motivational factor.

SMITH, CHARLES A., AND JENSEN, MYRTLE R. "Educational, Psychological, and Physiological Factors in Reading Readiness I, *The Elementary School Journal*, XXXVI (April, 1936), pp. 583.

The first part of a discussion of reading readiness: the problem, present practice, and psychological factors. Bibliography.

WALCOTT, FRED G. "New Methods and Objectives in Teaching Dull-normal Pupils to Read," *The School Review*, May, 1936), pp. 348-362.

The author presents the result of two experimental studies of the teaching of reading to pupils of normal and less than normal ability from which he concludes that our present theories and practices will need to be frequently modified in effective classroom instruction of pupils of this type.

WALKER, HELEN M. "Needed Improvement in the Teaching of Statistics," *Teachers College Record* (April, 1936), pp. 607-618.

The author discusses the planning of statistical courses for students at various levels; improving the teaching: statistical vocabulary content, clarification of concepts, laboratory work, etc.

NEW TESTS

ELWELL, FAYETTE H., AND FOWLKES, JOHN GUY. *Instructional Tests in Bookkeeping* (World Book Co., 1934), 91 pp.

The authors present a series of eighty tests to be employed in the teaching of bookkeeping at the high school and college levels.

LEITER, RUSSELL G., AND PORTEUS, STANLEY D. *The Leiter International Performance Scale*, University of Hawaii Research Publications, No. 13 (Honolulu: University of Hawaii, 1936), 42 pp.

The authors present a description of and data relative to the Leiter International Performance Scale. One of its most distinctive features is the fact that neither verbal instructions nor pantomime are necessary in applying the tests. It has been standardized for different cultural groups.

THURSTONE, L. L., AND THURSTONE, THELMA G. "The 1935 Psychological Examination," *The Educational Record*, (April, 1936), pp. 296-318.

A tabulation of the results from the 1935 edition of the American Council Psychological Examination for 266 colleges and 58,402 students.

TORGERSON, T. L., AND OTHERS. *The Wisconsin Plane Geometry Test, the Wisconsin Algebra Test, the Wisconsin Biology Test, the Wisconsin General Science Test, the Wisconsin American History, and the Wisconsin Language Usage Test*, (E. M. Hale and Co., 1936).

A series of achievement tests employed as a part of the statewide testing program carried on jointly by the School of Education, the Department of Mathematics of the University of Wisconsin and the Wisconsin State Department of Public Instruction.

RESEARCH NEWS and COMMUNICATIONS

Address all research news and communications to Carter V. Good, Teachers College, University of Cincinnati, Cincinnati, Ohio.

Vineland Research.—The research program at the Training School, Vineland, New Jersey, during 1934-35 has been in three major directions, namely, (1) the development of a scale for the measurement of social maturity, (2) the organization of a group of studies in the field of physiological psychology, and (3) the continuing evaluation of clinical case studies and case work. The director of research is Edgar A. Doll.

Reorganization at Parsons, Kansas.—The public schools of Parsons, Kansas, are this year operating on what is known as the "six-four-four" plan of organization. The new step in organization was taken at the opening of school in September, 1935, after the administrative staff under Superintendent Rees H. Hughes had been looking toward it for several years and after the board of education early in 1935 had given formal approval to the plan. The system for a number of years had six-year elementary schools, three-grade junior high schools, a three-grade senior high school, and a two-year junior college, which was housed with the senior high school. The six-year elementary schools remain undisturbed by the new arrangement. The two junior units now include Grades VII-X, and the former senior high school and junior college are now combined into a single junior-college unit including Grades XI-XIV. The new plan was instituted to afford a better service to the youth of the city through a simplification of organization, a strengthening of both junior high school and junior-college units, an improved continuity of the curriculum, and improved articulation within the whole system. In putting the reorganization into effect, the local school authorities are being advised by Leonard V. Koos of the University of Chicago, long a student and advocate of both junior high school and the junior college.

Wisconsin Studies.—A series of studies is being made of the fiscal, administrative, and attendance units of the system of public education in Wisconsin, by J. G. Fowlkes of the University of Wisconsin. The study of fiscal units includes an evaluation of the public school equalization fund law under which approximately 83 percent of all public school funds are distributed.

The University of Wisconsin High School, of which H. H. Ryan is principal, is one of the thirty secondary schools carrying on experimentation under the auspices of the Progressive Education Association. An experimental curriculum has been introduced to include one group of twenty pupils from each of the three upper classes. In all, eight such groups are to be started in the experimental curriculum. The project is set up in such a way that it will be possible to relate the curriculum to current results as well as to results after graduation from high school.

Records of the high school graduates are being kept in such a way as to throw light upon the validity of various items for prediction of college success.

A member of the high school staff is working under a subvention from the North Central Association in search of an answer to the question: "Can an adequate body of authentic material be found for use in the secondary school for the purpose of showing the actual working of our government—national, state, and local?"

Another member of the high school staff is working under a subvention from the North Central Association in the introduction of an experimental course in health in a number of high schools. A part of the plan is the evaluation of results, particularly with respect to the health of the pupils who are enrolled in the curriculum.

On Intelligence Tests at the College Level.—For a number of years most of the freshmen at Bethany College have been tested by two different intelligence tests. Because the data from these tests may be of interest to others who are working on testing problems at this level, there are listed, in the table below, the coefficients of correlation between the tests, the probable errors of the coefficients, and the means and standard deviations for each test. Unfortunately the number of cases is not large.

The tests included are: the American Council on Education Psychological Examination by L. L. and T. G. Thurstone; the Ohio State University Psychological Examination by H. A. Toops; and the Otis Self-Administering Tests of Mental Ability, Higher Examination (Form A) by A. S. Otis. The results reported here are for 1930 to 1935 inclusive.

This information was provided by Forrest H. Kirkpatrick of Bethany College.

Secondary School Standards. — More than seven hundred collaborators in all parts of the country have coöperated in the evaluation and constructive criticism of checklists developed during recent months by the Committee on the Coöperative Study of Secondary School Standards. These educators include the principals of public high schools of various types and sizes in every state, headmasters of many private and denominational secondary schools, representatives of state departments of education, and professors of education in leading schools of education.

Since the executive office of the Coöperative Study was opened in Washington in September, 1935, the Committee has formulated statements of guiding

TABLE I

COEFFICIENTS OF CORRELATION, PROBABLE ERRORS OF THE COEFFICIENTS, NUMBER OF CASES INVOLVED IN EACH CORRELATION, AND MEANS AND SIGMAS FOR INTELLIGENCE TESTS ADMINISTERED TO BETHANY COLLEGE FRESHMEN

X Variable	M X	S. D. X	Y Variable	M Y	S. D. Y	N	r	P.E. r
Otis S. A.	45.33	8.26	Thurstone 1935	192.5	59.8	96	.75	.0301
O. S. U. 18	81.73	22.9	Thurstone 1935	190.77	59.6	104	.84	.018
O. S. U. 18	81.84	23.45	Otis S. A.	45.16	8.24	98	.75	.0296
O. S. U. 18	80.816	22.3	Thurstone 1934	179.415	49.2	141	.79	.045
O. S. U. 18	80.382	25.1	Thurstone 1933	153.596	69.5	89	.846	.02
O. S. U. 17	118.775	38.97	Thurstone 1932	173.03	56.94	93	.87	.016
O. S. U. 17	105.91	36.9	Thurstone 1931	149.32	48.45	107	.78	.025
O. S. U. 16	139.15	41.325	Thurstone 1930	141.258	50.18	127	.796	.021

principles and tentative checklists covering almost 1400 items classified in five general areas. These have been distributed to the collaborators in order to secure their judgments and constructive suggestions for improvement. The five areas selected and the number of items in each are: the secondary school pupil, including articulation, admission, guidance, and records, 267 items; the secondary school staff, including qualifications, improvement, and conditions of service, 220 items; the educational program, including curriculum, pupil activities, library, supervision, and instruction, 562 items; and secondary school administration, including organization, business management, and community relations, 105 items.

The judgment of the collaborators who have studied these checklists is generally very favorable, but they have made hundreds of suggestions for modification and improvement before they are willing to approve them as entirely satisfactory instruments for purposes of evaluation, stimulation, and accreditation. Most of the judges seem to feel, however, that these proposed instruments, even in their present tentative form, look toward much more flexible, valid, and stimulating standards than those now in use in most parts of the country.

A series of six meetings of the General Committee of 21 members representing the six regional associations of colleges and secondary schools of the country was held in St. Louis the week preceding the meetings of the Department of Superintendence of the National Education Association. A number of leading educators met with the Committee to aid in evaluating the checklists and other material prepared to date, and to plan for the future activities of the Study.

Research Division, N.E.A.—"The evidence clearly indicates that during the present school year of 1935-36, progress has been made toward restoration and readjustment of salaries . . . It is important that the present period of salary adjustment be used to think through the whole problem of teachers' salaries. A mere return to 1929 conditions may not be adequate or equitable in 1936." This statement from the foreword to the January and March, 1936, *Research Bulletin* of the National Education Association indicates the necessity for careful study of the entire question of equitable salaries. Part one of "The Preparation of Teachers' Salary Schedules" deals with administrative and fact-finding procedures. Part two is concerned with the problem of drafting the schedule.

The question of an adequate standard of living for teachers is closely related to that of equitable salaries. In 1932 the Committee on the Economic Status of the Teacher authorized a study of cost-of-living data furnished by teachers in twelve monthly budget statements. A summary of the findings from 2,566 usable records was published as the September, 1935, *Research Bulletin* entitled "The Teacher's Economic Position."

The November, 1935, *Research Bulletin* on "Better Reading Instruction" is a survey of research and also of successful practice. Coöperation was obtained from two hundred three schools or school systems in conducting a questionnaire study of successful teaching of elementary school reading.

The Research Division also assisted in the preparation of the 1935 reports of the Committee on Tenure for Teachers and the Committee on Retirement Allowances of the National Education Association: "Recent Court Decisions on Teacher Tenure," "Teacher Tenure Legislation in 1935 to Date," and "Teacher Retirement Legislation in 1934 and 1935 to Date."

The Research Division has also continued to coöperate in an editorial and consultative capacity with regard to the publications of certain departments of the National Education Association. *The Social Studies Curriculum*, the 1936 Yearbook of the Department of Superintendence, is a practical handbook dealing with the principles and practices of a modern program of social studies. The material is divided into three parts: factors conditioning the social studies; the social studies curriculum; and teaching, evaluating and revising the social studies program. Among the factors conditioning the social studies are: the nature of society; the role of education and the school; and the nature of the learner and the learning process. Recognition that a social studies program for the schools is a compromise between the comprehensiveness and specialism of scholarship and the requirements and necessities of schoolroom organization and presentation is basic to successful teaching.

The Fourteenth Yearbook of the Department of Elementary School Principals entitled *Socializing Experiences in the Elementary School* was issued in July, 1935. The socializing influence of the school may be increased by modifying the traditional content and methods of teaching in the regular subjects, and by the introduction or adaptation of a wide variety of life-like group activities, such as clubs, assemblies, publications, and pupil participation in student management. This yearbook is primarily concerned with these more informal, life-like, and formerly "extra" activities, including also chapters on the activity program and coöperative community activities.

Current issues of the *Review of Educational Research* published by the American Educational Research Association comprise these titles: 1. "Psychological Tests," Vol. V, No. 3, June, 1935. 2. "The School Plant," Vol. V, No. 4, October, 1935. 3. "Educational Tests and Their Uses," Vol. 5, No. 5, December, 1935. 4. "Mental and Physical Development," Vol. VI, No. 1, February, 1936. 5. "Pupil Personnel, Guidance, and Counseling," Vol. VI, No. 2, April, 1936.

The Educational Research Service, maintained by the Research Division and the Department of Superintendence, has issued these circulars since June, 1935:

1. "Education Discussed in Lay Magazines," Circulars 7, 9, and 10 (1935); 1 and 4 (1936). These five summaries of articles in non-educational magazines cover the period from April 1, 1935, to April 1, 1936.

2. "Questionnaire Studies Completed —Bibliography No. 6, 1934-35," Circular No. 8, September, 1935. The list of thirty-two studies presented in this circular is based on questionnaires submitted to city and state superintendents of schools.

3. "School Expense Compared with Combined City and School Expenses, 1934," Circular No. 2, February, 1936. This circular presents information for the ninety-four largest cities on total city and present school expense, school expense only, cost per capita, and percent the school cost is of total expense.

4. "Opening and Closing Dates of City School Terms, 1935-1936; Dates of College Summer Sessions, 1936," Circular No. 3, March, 1936. The facts in this circular are based on an inquiry sent to the three largest school systems in each state and the three largest colleges and universities in each state which report the largest enrolments in summer school.

Office of Education Projects.—President Roosevelt has authorized the Office

of Education to conduct five projects for the advancement of education, financed by relief funds and designed to give work to some 3500 "white collar" unemployed:

(1) Forums: Authorization for this project ($330,000) will permit establishment of demonstration centers in various parts of the country to develop further and test the plan for civic education now in its fourth year at Des Moines. These public forums (ten in all) will be administered entirely by local groups, but the lessons learned locally will be passed on to other communities using this approach to adult education. The Director is Commissioner Studebaker.

(2) Research: Under this project needy graduates will engage in coöperative studies of important educational problems. Results will be brought together by the Office of Education for use throughout the nation. Appropriation: $500,000. Director: Ben W. Frazier of the Division of Higher Education.

(3) Vocational Education for Negroes: There is a great need for facts to be used as a basis for improvement of Negro education. Studies to be conducted as part of this project will give a better picture of the entire problem of educational and vocational opporunities for Negroes than we have ever had. Appropriation: $234,000. Director: Ambrose Caliver, specialist in the education of Negroes.

(4) Radio: For the first time a truly adequate staff will be made available through this project to present a series of educational radio programs. Talent selected from CCC and WPA dramatic rolls will staff an educational radio "workshop" in Washington, D. C. Appropriation: $75,000. Director: William Dow Boutwell, Chief of the Editorial Division.

(5) Organization of Local School Units: This project, designed to aid states in planning for more efficient school district organization, was originally submitted to include 31 states. As approved it now includes only 10 states. Appropriation: $844,000. Director: Henry F. Alves, specialist in state school administration.

Assistant Commissioner Bess Goodykoontz is in general charge of projects Nos. 2, 3, and 5.

Federal relief funds totaling $1,983,000 are available for the five projects.

Visual Instruction.—Questionnaires for the National Visual Instruction Survey have been sent to 21,000 city and county superintendents of schools and principals of private high schools throughout the United States from the Office of Education by Cline M. Koon, Specialist in Radio and Visual Education, and Director of the Survey. This survey of visual aids used in public and private schools, first of its kind ever undertaken on a national scale, is a coöperative endeavor of the American Council on Education and the United States Office of Education. Among other agencies expressing approval of the inquiry are the National Education Association, National Congress of Parents and Teachers, Rural Electrification Administration, and Division of Motion Pictures, U. S. Department of the Interior.

Superintendents and principals of schools quite generally report insufficient budgetary provisions for visual aids and insufficient training of teachers in their use. Most visual instruction in schools today, according to the first few hundred questionnaires returned, is in geography, history, health, travel, and safety education, although superintendents and principals report using vis-

ual aids for teaching economics, English, guidance, industry, literature, biology, music, agriculture, reading, crafts and drama. Use of news reels, feature pictures, foreign language films, and religious education visual aids were also reported.

The study will also bring information to the U. S. Office of Education for the first time on the use of radios, phonographs, and centralized radio-sound equipment in public and private schools. Suggestions were invited on how national agencies can facilitate use of visual aids for instructional purposes. The replies will supply information on the nature and use of audio-visual aids, and present visual-instruction needs of school systems.

Vocational Education in Alaska.—The Department of Education of Alaska was established nineteen years ago, and of course is quite young as compared with other state or territorial school systems. Partly due to its youth and partly because of the few people employed in the department, it has not as yet a department of research and, hence, no scientific studies have been made relative to educational needs.

However, those who have been connected with Alaska's educational system for any length of time have seen the need of an enlargement of its vocational program. Prior to the present year, work has been offered in some of the larger schools in manual training and home economics, but this has been of the old type and without federal aid.

In April, 1934, Congress made the provisions of the George-Ellzey Act applicable to Alaska, providing $15,000 annually for aid in vocational education whenever the Territory accepted the provisions of the Act and appropriated a like amount. The 1935 Legislature of Alaska accepted the provisions; plans were immediately draw up by the Department, and these were approved by the Office of Education of the Federal Government.

One supervisor was employed for the three phases of work, namely, home economics, agriculture, and trades and industries, with a part-time assistant supervisor of home ecenomics. The supervisor started work in July, 1935, visiting school boards and superintendents in all the larger cities.

Due to the fact that teachers had already been employed for home economics and manual training in all the schools, not all schools are offering the vocational work this year because some of the teachers so selected could not be approved under the adopted plans. However, school boards now know the requirements, and qualified teachers will be selected hereafter.

A good start has been made, however, in adult education in nearly all the larger cities. Courses have been offered in mining, marine engineering, navigation, taxidermy, cooking, sewing, etc. Alaska is so rich in its many natural resources that it presents a splendid opportunity for instruction along these lines.

The Smith-Hughes Act has never applied to Alaska, which is the only political subdivision of the United States so situated. A bill is now before Congress providing for the application of this Act to Alaska, and there seems every reason to believe that it will pass. This will give the sum of $15,000 more from the Federal Government, and with that appropriated by the Territorial Legislature, will make a total of $60,000 available for vocational work.

When this sum is available, special supervisors will be appointed in each of the three fields of vocational education, and instruction will be extended to the

older pupils of as many of the seventy-five village schools as funds will permit.

In these rural schools, up to the present time, only the usual academic work has been offered. Many of these schools are quite isolated and few of the parents are financially able to send pupils to high school, so few ever go farther than the eighth grade.

The plan is to start the vocational work in the sixth or seventh grade and to give two or three years of instruction along lines which will better fit the girls for the work they usually do and offer the boys training in the type of work which they will follow in their respective communities. At the same time, the academic work will still be given, so that those pupils who can afford to go farther in school work will be adequately prepared.

Such a program, it is believed, will be of lasting economic benefit to the pupils, as well as to the Territory.

This information was provided by Anthony E. Karnes, Commissioner of Education, Territory of Alaska.

Improvement of Instruction in Arkansas.—The Arkansas Coöperative Program to Improve Instruction is now in the third year of a five-year plan. Over 25,000 teachers and patrons are actually participating. The program was initiated by the State Department of Education and is being directed by a member of the Department staff. All teacher-training institutions and other educational agencies of the state are coöperating. Doak S. Campbell and Hollis L. Caswell of George Peabody College for Teachers are acting as general consultants. Special consultants and laboratory assistants from various sections of the country have been secured to direct certain phases of the work.

Arkansas has 460,000 school children under the guidance of over 12,000 teachers. The general purpose of this program is to improve the daily program of school life of these pupils and teachers. It is proposed through voluntary coöperation and extended participation, including all local, county, regional, and state educational agencies and institutions, to initiate, stimulate, and guide group study and classroom experiments in curriculum revision. It is not proposed in any way to minimize the fundamental role of "organized subject matter," but rather to vitalize subject-matter and other instructional materials by tying them into the experiences, interests, and realized needs of the pupils. When the purpose of the program is realized, logically arranged subject matter will no longer serve as a basis for organizing instruction.

After establishing an administrative organization designed to utilize all available educational agencies, summer curriculum laboratories were set up at Peabody College and the University of Arkansas. Selected leaders under the direction of general and special consultants produced bulletin materials for guidance of teachers and patrons. These "curriculum laboratories" have been conducted each summer. Over 200 selected teachers have been trained in the philosophy, production, and organization of new curriculum materials. More than 40,000 copies of these bulletins have been produced and are now in use throughout the state. Tentative principles, aims, scope, and study materials have been set forth. Thousands of teachers and patrons in scores of local, county, regional, and state groups have studied these materials. A new instructional program based on this experimental point of view is gradually coming into being. Teachers, pupils, and patrons, from the larger cities to the

one-room log school houses in the mountains, have "come alive" under the stimulation of this coöperative experimental enterprise. With direction from local and state supervision, hundreds of classrooms have become "little laboratories." Teachers in ever-increasing numbers are shifting from the old page assignment to the new larger units of instruction centered in the interests, experiences, and needs of the children. Scores of these child-centered classrooms are going into action. This fine theory is gradually becoming a practical reality in "key schools" in all sections of the state.

The Arkansas Congress of Parents and Teachers is making an outstanding contribution. They have produced 8,000 copies of two bulletins and have enrolled 10,000 persons in study, discussion, and forum groups. Their membership has increased from fifteen to thirty thousand during this two-year period. More than a score of other social and professional clubs are coöperating this year in a study of "Changing Attitudes through Adult Education." Other special features are: (1) special curriculum courses are being offered in all approved colleges in the state; (2) experimental and demonstration teaching is offered in the summer practice schools of these institutions; (3) seventy-five percent of the 2,000 Negro teachers in the state are actively participating; (4) key schools nave been designated. Under special supervision, some of the many problems constantly coming to the surface are being attacked first by these teachers and schools. No effort is made to hold all schools in the same phase. While many schools are in the initial steps, others have a well organized program under way. On the whole, elementary schools develop more rapidly than secondary schools.

To summarize briefly, an experienced-centered curriculum, embracing certain principles, aims, and scope is held as a tentative hypothesis to guide Arkansas teachers in applying the philosophy of experimentalism to the instructional program of more than 12,000 classrooms. Purposeful activity is supplanting arbitrary intellectualism. Teachers are getting away from the idea that it is so necessary to *give* pupils something; they are learning to center the instructional program *within the interests and needs of the children*, to accept and to have faith in the individual capacity for growth through natural development. Thousands of trained adults and tens of thousands of school children are enjoying this experiment in improving school life. Many problems are yet unsolved, even undiscovered, but a newer, better instructional program is being wrought out and fashioned to meet the changing needs of the school children of Arkansas.

This information was provided by Ed. McCuistion, Director of Certification and Teacher Training.

Research in Australia, 1934-35.—The activities of the Australian Council for Educational Research may be summarized as follows:

1. The Council is the only permanent educational establishment in Australia devoted to the purely professional interests of education and embracing all States. It is generally recognized that the greatest need in Australian education to-day is the development of a greater professional spirit and status. This requires that education should be a progressive study. A very substantial contribution towards this end can be made by an organization devoted to the study of educational trends in various parts of the world and having the machinery for

supporting inquiry and for initiating experiments.

2. Being quite independent of existing organizations and having no administrative functions the Council is able to express impartial opinions on educational problems. The fact that the Council has the active support of almost all leading educationalists in Australia lends weight to those opinions.

3. During its five years it has supported over one hundred investigations and will shortly have published forty reports. Many of these reports have an immediate bearing on the practical problems of the classroom teacher. Their publication cannot but have a marked influence on Australian education.

4. The Council is in a position to sponsor surveys and inquiries. Special mention must be made of the survey of library conditions throughout Australia—a survey which is destined to be of historical importance in library matters and to have important reactions upon the work of the schools—and of the inquiry into the problems of secondary education which has just been opened.

5. The Council is making good progress in supplying standardized tests of attainment in the chief school subjects as well as tests of general intelligence. Australia is many years behind in such matters. The existence of such tests enables any teacher to ascertain precisely the standards of his pupils and removes one of the strongest arguments against the view that school should do most of its own examining.

6. The Council acts as a clearing house of information on educational developments. Through the action of the International Institute of Intellectual Coöperation the Council is recognized as the national center in Australia for the provision of educational information. Exchanges of information and of publications are maintained between the Council and a number of similar centers in other parts of the world. The status of the Council in these matters is recognized by officials of the Commonwealth Government and of the State Education Departments.

7. Owing to its relationships with overseas centers the Council is able to assist and advise teachers and administrators going abroad. It has also been instrumental in arranging for a number of Australian educationalists to receive grants for traveling purposes from the Carnegie Corporation of New York.

8. The Council maintains a library of educational books which is freely available to all teachers. Some of these books are not available in other libraries in Australia. Special attention is paid to expensive books of a technical character. At present the library contains about 800 volumes, of which approximately one-eighth are on loan to teachers at any given time.

9. The chief educational journals published in the English language are received and a card-index record is kept of important articles appearing in them. This makes it possible to provide help to a number of inquirers who wish to write theses, prepare papers, or give addresses at teachers' meetings and conferences.

10. The Council has in hand arrangements for a conference on education in Australia in 1937 under the auspices of the New Education Fellowship (England). This conference, with leading speakers from overseas, is likely to be the most important event of the kind which has so far taken place in Australia.

11. Among the practical matters which the Council or its workers have investigated since its inception may be mentioned: the curriculum of the primary

school; the use of tests for predicting school success; defects of speech in school children; color blindness; the absorptive capacity of various occupations; the length of the teacher's professional life; the best methods of revision of school work; juvenile delinquency; the value of school broadcasting; the scholarship system in Victoria (to be published shortly); the teaching of history and civics; class groupings; and education of retarded children.

However, the Council has not sought arbitrarily to centralize or to control research. It is a scientific society, working on a coöperative basis with state educational authorities, but independent of them. The investigations are conducted by individuals and committees whose efforts are co-ordinated and supported by the Council.

The Council planned to test in March, 1936, about 30,000 children in Australia between the ages of nine and fourteen. One of the tests used was the Otis Self-Administering Test of Intelligence. The New Zealand Council for Educational Research has planned a similar program of measurement. Scotland has tested practically every child in its territory between 10½ and 11½ years of age.

Stanford Studies.—P. E. Davidson and H. D. Anderson have completed a study of the "Occupational Mobility in an American Community," in which the characteristics of a seven percent sample of the male working population of San Jose, California, were analyzed. The findings relate to degree of occupational inheritance and mobility, juvenile jobs, entrance to permanent occupation, individual movement in occupational status thereafter, schooling, income, marital conditions, dependency, displacement, unemployment, etc.

G. N. Kefauver and H. C. Hand have completed "An Appraisal of Guidance in Secondary Schools" which is to be published by Macmillan during the current year. Different aspects of the development of pupils in ten cities were measured. The findings show variation in the status of pupils from city to city, with a relationship to the emphasis in the guidance program. The status of pupils and the guidance program were not satisfactory in all respects for any of the cities studied.

Comprehensive Examinations in the Chicago Junior Colleges.—Courses in English, composition, biological science, the humanities, physical science, and social science constitute the required portion of the curriculum of the three Chicago municipal junior colleges. The typical freshman studies English, composition, the first year of social science, biological science or physical science, and a number of elective courses. The typical sophomore enrolls in the humanities, biological science or physical science, the second year of social science, and several elective courses. The achievements of the students enrolled in the required courses named above are measured by comprehensive examinations prepared by the Department of Examinations, with the assistance of the faculties concerned. Each of these examinations is designed to cover one full year of study. The scores and marks are used in determining eligibility for graduation. They are also used to provide data on which to base recommendations of students to higher institutions, and to provide information useful in educational and vocational counseling. Ultimately, six comprehensive examinations, corresponding to the courses named, will be given each semester. No examination for the second-year

course in social science was necessary in June, 1935.

The following topics are considered in a report describing the 1935 examinations, prepared by M. D. Engelhart: (1) the preparation of the examinations; (2) the administration; (3) the scoring; (4) the D-score; (5) the assignment of letter grades; (6) the distributions of original scores, D-scores, and letter grades; (7) the reliability of the examinations; and (8) the analysis of the achievement of the "average" student. The report concludes with some inferences derived from the data and with some recommendations based upon these inferences.

Iowa Testing Program.—For a long time a number of educational workers have felt the need of types of appraisal of instruction which give an adequate picture of what is being accomplished, which show deficiencies wherever they may exist, and which afford data that may readily be made the basis for a program of improvement. It is claimed that, under the direction of E. F. Lindquist and with the coöperation of the staff of the College of Education, University of Iowa, such a battery of tests has been developed.

For the past eight years the College of Education and the Extension Division have provided a coöperative regional testing service known as the Every-Pupil Testing Program. One of the most important outgrowths of this experience has been the development of a comprehensive battery of tests of the basic skills in reading, study, language, and arithmetic for grades six to eight. The second (1936) edition of these tests was given to over 40,000 students in January, 1936.

It has been decided to release the tests for nation-wide use, independent of the organized program, after norms have been established on the basis of the results of that program.

It is said that the whole project is being conducted on a strictly non-commercial basis with no royalties being paid. All proceeds from the sale of the tests will be devoted to research and experimentation for the future production of improved materials and service.